

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:36:57 ; Search time 43 Seconds
(without alignments)
6072.563 Million cell updates/sec

Title: US-09-556-246-1
Perfect score: 7526
Sequence: 1 MAMKTLPIYLLLLSVFVIQ.....ARAITRSGQTLSKWNCP 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues
Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	US-10-124-557-2	Sequence 2, Appl1
2	7526	100.0	1404	US-10-124-557-62	Sequence 62, Appl1
3	7291.5	96.9	1361	US-10-124-557-40	Sequence 40, Appl1
4	7245.5	96.3	1363	US-10-124-557-52	Sequence 52, Appl1
5	7232	96.1	1354	US-10-124-557-48	Sequence 48, Appl1
6	7011	93.2	1330	US-10-124-557-46	Sequence 46, Appl1
7	7011	93.2	1330	US-10-124-557-60	Sequence 60, Appl1
8	7007.5	93.1	1311	US-10-124-557-42	Sequence 42, Appl1
9	6971	92.6	1314	US-10-124-557-50	Sequence 50, Appl1
10	6951.5	92.4	1313	US-10-124-557-142	Sequence 142, Appl1
11	6971	89.4	1270	US-10-124-557-44	Sequence 44, Appl1
12	6085	80.9	1140	US-10-124-557-104	Sequence 104, Appl1
13	5510.5	73.2	1049	US-10-124-557-58	Sequence 58, Appl1
14	5425.5	72.1	1022	US-10-124-557-84	Sequence 84, Appl1
15	5335	70.9	1038	US-10-124-557-74	Sequence 74, Appl1

16	5011	66.6	941	14	US-10-124-557-14	Sequence 14, Appl1
17	2850.5	37.9	538	12	US-10-038-694-3	Sequence 3, Appl1
18	2034.5	27.0	463	14	US-10-124-557-54	Sequence 54, Appl1
19	1763.5	23.4	423	14	US-10-124-557-66	Sequence 66, Appl1
20	1754	23.3	422	14	US-10-124-557-68	Sequence 64, Appl1
21	1460	19.4	372	14	US-10-124-557-64	Sequence 96, Appl1
22	1188	15.8	220	14	US-10-124-557-96	Sequence 106, Appl1
23	1188	15.8	5179	9	US-09-922-217-1068	Sequence 1068, Appl1
24	1188	15.8	5179	10	US-09-833-263-1068	Sequence 1068, Appl1
25	1188	15.8	5179	14	US-10-023-380-1068	Sequence 94, Appl1
26	1126	15.0	209	14	US-10-124-557-94	Sequence 116, Appl1
27	1121	14.9	208	14	US-10-124-557-132	Sequence 92, Appl1
28	1111.5	14.8	207	14	US-10-124-557-136	Sequence 90, Appl1
29	1111.5	14.8	207	14	US-10-124-557-136	Sequence 116, Appl1
30	1100	14.6	204	14	US-10-124-557-92	Sequence 92, Appl1
31	1077	14.3	296	14	US-10-124-557-70	Sequence 70, Appl1
32	1074	14.3	231	14	US-10-124-557-30	Sequence 30, Appl1
33	1041	13.8	192	14	US-10-124-557-90	Sequence 90, Appl1
34	968.5	12.9	188	12	US-10-038-694-2	Sequence 2, Appl1
35	945	12.6	172	14	US-10-124-557-88	Sequence 88, Appl1
36	881	11.7	188	14	US-10-124-557-32	Sequence 32, Appl1
37	871	11.6	156	14	US-10-124-557-106	Sequence 106, Appl1
38	871	11.6	156	14	US-10-124-557-118	Sequence 118, Appl1
39	827.5	11.0	157	14	US-10-124-557-118	Sequence 102, Appl1
40	827.5	11.0	157	14	US-10-124-557-114	Sequence 114, Appl1
41	809	10.7	1367	10	US-09-801-368-108	Sequence 108, Appl1
42	806	10.7	217	14	US-10-124-557-76	Sequence 76, Appl1
43	805	10.7	1325	9	US-09-864-761-35612	Sequence 35612, A
44	800	10.6	141	14	US-10-124-557-80	Sequence 80, Appl1
45	750	10.0	131	14	US-10-124-557-82	Sequence 82, Appl1

ALIGNMENTS

RESULT 1
US-10-124-557-2
Sequence 2, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:

NAME: Cseir, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-10-124-557-2

Query Match 100.0%; Score 7526; DB 14; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLILLSVFIYQVSSQDLSGACGCGEGRSDATTCNCDYNCQHYMECCDF 60
DB 1 MAMKTLPIYLILLSVFIYQVSSQDLSGACGCGEGRSDATTCNCDYNCQHYMECCDF 60
QY 61 KRVTAEISCKRCESFERGECDCDAQCKKYDKCCPDYESFCAEVNPTSPPSSKAP 120
DB 61 KRVTAEISCKRCESFERGECDCDAQCKKYDKCCPDYESFCAEVNPTSPPSSKAP 120
QY 121 PPSGASQITKSTTKSPKPPNKKTKYVIESEITEEHSVSENQSSSSSSSSSSSTTW 180
DB 121 PPSGASQITKSTTKSPKPPNKKTKYVIESEITEEHSVSENQSSSSSSSSSSSTTW 180
QY 181 KIKSSKNAANRELOKLVKQDNKKRKKKPTKPPVVDGSGLDGNDGFVTTTDPST 240
DB 181 KIKSSKNAANRELOKLVKQDNKKRKKKPTKPPVVDGSGLDGNDGFVTTTDPST 240
QY 241 TOHNKVSSTPKITTAKPINRPSLPNSDTSKETSITVNKETVTEKETTNNKQSTDG 300
DB 241 TOHNKVSSTPKITTAKPINRPSLPNSDTSKETSITVNKETVTEKETTNNKQSTDG 300
QY 301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTGKALTTPEKPTTPEKPA 360
DB 301 KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTGKALTTPEKPTTPEKPA 360
QY 361 TTPKEPTTTKSAPTTPEKPAPTTTSKAPPTTPEKPAPTTPEKPAPTTPEKPA 420
DB 361 TTPKEPTTTKSAPTTPEKPAPTTTSKAPPTTPEKPAPTTPEKPAPTTPEKPA 420
QY 421 APTTTSKAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 480
DB 421 APTTTSKAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 480
QY 481 BPAPAPKAPAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 540
DB 481 BPAPAPKAPAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 540
QY 541 TTKSAPTTPEKSPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 600
DB 541 TTKSAPTTPEKSPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 600
QY 601 APTAKEPAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 660
DB 601 APTAKEPAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 660
QY 661 PEEBAPPTTPEKAAAPNTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 720
DB 661 PEEBAPPTTPEKAAAPNTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPA 720
QY 721 APTTCKAPKELATTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEKPA 780
DB 721 APTTCKAPKELATTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEKPA 780
QY 781 TAPTLKEBAPPTTPEKPAKELAPTTTGGPTSTSDKAPPTTPEKPAPTTPEKPA 840
DB 781 TAPTLKEBAPPTTPEKPAKELAPTTTGGPTSTSDKAPPTTPEKPAPTTPEKPA 840

```

```

DB 781 TAPTLKEBAPPTTPEKPAKELAPTTTGGPTSTSDKAPPTTPEKPAPTTPEKPA 840
QY 841 KPAPTTPEPTTSEVSTPTTKEPTTIKSPDESIPBELSAETPALENSPEBQVPT 900
DB 841 KPAPTTPEPTTSEVSTPTTKEPTTIKSPDESIPBELSAETPALENSPEBQVPT 900
QY 901 TKTPAATPEWTTAKKTERDRLTTPETTTAPAKMTKETATTEKTESKIRATTTQV 960
DB 901 TKTPAATPEWTTAKKTERDRLTTPETTTAPAKMTKETATTEKTESKIRATTTQV 960
QY 961 TSITTDTPPKITTLTKITTLAPRVYTTKTIITTEIMNKEETAKPKDRATNSKATTPK 1020
DB 961 TSITTDTPPKITTLTKITTLAPRVYTTKTIITTEIMNKEETAKPKDRATNSKATTPK 1020
QY 1021 POKTKAPKPTSTKTKRTPRVAKPRTTTPRKGSTMBELNPTSIAEAMQTTTRPN 1080
DB 1021 POKTKAPKPTSTKTKRTPRVAKPRTTTPRKGSTMBELNPTSIAEAMQTTTRPN 1080
QY 1081 QTPNSKLYEVNPKSEDGAGETPHMLLRPHVMPVTPMDYLPVPNOGIIINPMLS 1140
DB 1081 QTPNSKLYEVNPKSEDGAGETPHMLLRPHVMPVTPMDYLPVPNOGIIINPMLS 1140
QY 1141 DETNICNGKPYDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGISPIDTVFT 1200
DB 1141 DETNICNGKPYDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGISPIDTVFT 1200
QY 1201 RCNCEGKTFPPKDSQYRFTNDIDAGYPRPIFNGFGGLTQIYAAASTAKYKWPSSVY 1260
DB 1201 RCNCEGKTFPPKDSQYRFTNDIDAGYPRPIFNGFGGLTQIYAAASTAKYKWPSSVY 1260
QY 1261 FFKRGGSIOQYIYQOEPPQKCPGRPALNYPVYEMTOVRRRREARIGSQHTIRIQY 1320
DB 1261 FFKRGGSIOQYIYQOEPPQKCPGRPALNYPVYEMTOVRRRREARIGSQHTIRIQY 1320
QY 1321 SPARLAYQDKGVLANEYKVSILMRGLPNVVTSAISLPIRKPDGYDYAASQDYINIDV 1380
DB 1321 SPARLAYQDKGVLANEYKVSILMRGLPNVVTSAISLPIRKPDGYDYAASQDYINIDV 1380
QY 1381 PSRTARAITTRSGOTLSKWYNCP 1404
DB 1381 PSRTARAITTRSGOTLSKWYNCP 1404

```

RESULT 2
 US-10-124-557-62
 ; Sequence 62, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesner, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gaert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: G1 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62

Query Match 100.0%; Score 7526; DB 14; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARKTPIYLLLLSVVVIQVSSQDLSGACGRCGEGYSDATCNCDYNOHMECCPDF 60
DB 1 MARKTPIYLLLLSVVVIQVSSQDLSGACGRCGEGYSDATCNCDYNOHMECCPDF 60
QY 61 KRVCTABLSCGRCFESFERGRCDCDAQCCKTKDKCCPDYSEFCAEVHNTPSPSSKAP 120
DB 61 KRVCTABLSCGRCFESFERGRCDCDAQCCKTKDKCCPDYSEFCAEVHNTPSPSSKAP 120
QY 121 PRGASQTIKSTTKRSPKPNKKTKKVISEBITTEHVSSENOESSSSSSSSSTIW 180
DB 121 PRGASQTIKSTTKRSPKPNKKTKKVISEBITTEHVSSENOESSSSSSSSSTIW 180
QY 121 PRGASQTIKSTTKRSPKPNKKTKKVISEBITTEHVSSENOESSSSSSSSSTIW 180
DB 121 PRGASQTIKSTTKRSPKPNKKTKKVISEBITTEHVSSENOESSSSSSSSSTIW 180
QY 181 KISSKNSAANRELQKCLKYKDNKKKRTKKKPKPPVNDGASGLDNGFKATTPDTST 240
DB 181 KISSKNSAANRELQKCLKYKDNKKKRTKKKPKPPVNDGASGLDNGFKATTPDTST 240
QY 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNEKETTVEKETTNNKOTSDG 300
DB 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNEKETTVEKETTNNKOTSDG 300
QY 301 KEKTSKAKETOSIEKTSANDLAPTSKVLAKPTKAEITTTGPALETTTPKEPTTPKEPAS 360
DB 301 KEKTSKAKETOSIEKTSANDLAPTSKVLAKPTKAEITTTGPALETTTPKEPTTPKEPAS 360
QY 361 TTPEKPTPTTIKAPPTPKBPAPTTTKSAPPTPKBPAPTTTKBPAPTTTKBP 420
DB 361 TTPEKPTPTTIKAPPTPKBPAPTTTKSAPPTPKBPAPTTTKBPAPTTTKBP 420
QY 421 APPTTSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 480
DB 421 APPTTSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 480
QY 481 EPAPIAPKKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAP 540
DB 481 EPAPIAPKKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAP 540
QY 541 TTTSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 600
DB 541 TTTSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 600
QY 601 APTAPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 660
DB 601 APTAPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 660
QY 661 PEEPAPTPKAAAPNTPEKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAP 720
DB 661 PEEPAPTPKAAAPNTPEKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAP 720

DB 661 PEEPAPTPKAAAPNTPEKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAP 720
QY 721 APTPKKAPKELAPTTPEPTSTSDKAPPTPKGATTTKEPAPTTPKBPAPTTPK 780
DB 721 APTPKKAPKELAPTTPEPTSTSDKAPPTPKGATTTKEPAPTTPKBPAPTTPK 780
QY 781 TAPTTKEPAPTTPKKAPKELAPTTKGPSTTSKBPAPTTPKBPAPTTPKBPAPTTPK 840
DB 781 TAPTTKEPAPTTPKKAPKELAPTTKGPSTTSKBPAPTTPKBPAPTTPKBPAPTTPK 840
QY 841 KAPPTPEPTPEPTSEVSTPTTKEPTTIHKS PDSPELSAEPKALENSKEGVP 900
DB 841 KAPPTPEPTPEPTSEVSTPTTKEPTTIHKS PDSPELSAEPKALENSKEGVP 900
QY 901 TKTPAATKEMTTAKOKTTERDLRPTPTTAAPMTKETATTTKTESKLTATTTQV 960
DB 901 TKTPAATKEMTTAKOKTTERDLRPTPTTAAPMTKETATTTKTESKLTATTTQV 960
QY 961 TSTTQDTTPFKITTLKTTTLAPKVTTKKTIITTEINMKPESTAKPKDRTNSKATTPK 1020
DB 961 TSTTQDTTPFKITTLKTTTLAPKVTTKKTIITTEINMKPESTAKPKDRTNSKATTPK 1020
QY 1021 POKPTAPKPKPTSTKPKKPMKPVKPKTTPRCKMTSTMPBELNPTSRIAEAMLQTTTBN 1080
DB 1021 POKPTAPKPKPTSTKPKKPMKPVKPKTTPRCKMTSTMPBELNPTSRIAEAMLQTTTBN 1080
QY 1081 QTPNSKLVVNPKSEBAGABESTPMLLRPHVFMPEVTPDMDYLRPVNOCIIINPMIS 1140
DB 1081 QTPNSKLVVNPKSEBAGABESTPMLLRPHVFMPEVTPDMDYLRPVNOCIIINPMIS 1140
QY 1141 DETNINCGRPVGLTTLRNGTLVAFRGHYFMMLSPPSPARRITEVNGISPIDTVP 1200
DB 1141 DETNINCGRPVGLTTLRNGTLVAFRGHYFMMLSPPSPARRITEVNGISPIDTVP 1200
QY 1201 RCNCEGKTEFFKDSQYWRFTNDIKDAGYKPIFKGFGJLTGQIVAAALSTAKYKMBESY 1260
DB 1201 RCNCEGKTEFFKDSQYWRFTNDIKDAGYKPIFKGFGJLTGQIVAAALSTAKYKMBESY 1260
QY 1261 FRKGGSIQOYIYKQEPVQKCGRRPALNYPVYGENTQVRRRRFEPAIEPSQHTTRIQ 1320
DB 1261 FRKGGSIQOYIYKQEPVQKCGRRPALNYPVYGENTQVRRRRFEPAIEPSQHTTRIQ 1320
QY 1321 SPARLAODKGLVHNEVKVSIIMRGLPNVTSAISLPNIRKPDGYDYAFSQQOYINIDV 1380
DB 1321 SPARLAODKGLVHNEVKVSIIMRGLPNVTSAISLPNIRKPDGYDYAFSQQOYINIDV 1380
QY 1381 PSTARAITTRSGQTLISKVYNCP 1404
DB 1381 PSTARAITTRSGQTLISKVYNCP 1404

RESULT 3
US-10-124-557-40
Sequence 40, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetide Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible


```

/ STREET: 87 Cambridgepark Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Releasee #1.0, Version #1.25
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/124,557
/ FILING DATE: 16-Apr-2002
/ CLASSIFICATION: <Unknown>
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ APPLICATION NUMBER: US 07/390,901
/
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Coeert, Lueann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/
/ INFORMATION FOR SEQ ID NO: 52:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1363 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match          96.3%; Score 7245.5; DB 14; Length 1363;
Best Local Similarity 97.1%; Pred. No. 0;
Matchee 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAMKTLPIYLILLLSVFVIOVSSODLSGACRCEGYSRDATCNDVNCQHYMECCPFP 60
DB 1 MAMKTLPIYLILLLSVFVIOVSSQ----- 25
QY 61 KKVCTAELSGCRCEPESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 120
DB 26 -----ELSGCRCEPESFERGREGCDCAQCKYDKCCPDYESFCAEVHNPTSPSSKAP 79
QY 121 PPSGASQITKSTTKSSPKRPNNKTKTKVIESEBITTEHSVSSNOESSSSSSSSSTIW 180
DB 80 PPSGASQITKSTTKSSPKRPNNKTKTKVIESEBITTEHSVSSNOESSSSSSSSSTIW 139
QY 181 KIKSGNSANRELQKLVKONKKNRTKKKPTPKRPVVDZAGSGLDNDDEKVTTPDST 240
DB 140 KIKSSNSANRELQKLVKONKKNRTKKKPTPKRPVVDZAGSGLDNDDEKVTTPDST 199
QY 241 TQHNKVSPTKITTAKPINSRPSLPNSDTSKETSILTUNKETTVETKETTTNNKQSTDG 300
DB 200 TQHNKVSPTKITTAKPINSRPSLPNSDTSKETSILTUNKETTVETKETTTNNKQSTDG 259
QY 301 KEKTTSAKETSIEKTSADOLAPTSKVLAKPTPKAETTTKGPALATTPKEPTPTPKEPAS 360
DB 260 KEKTTSAKETSIEKTSADOLAPTSKVLAKPTPKAETTTKGPALATTPKEPTPTPKEPAS 319
QY 361 TTPKEPTPTTISAPTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
DB 320 TTPKEPTPTTISAPTTKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 379
QY 421 APTTKSAPTTKEPAPTTPKKAPATTPKEPAPTTPKEPTPTTKEPAPTTKEPAPTTPK 480

```

```

DB 360 APTTKSAPTTKEPAPTTPKKAPATTPKEPAPTTPKEPTPTTKEPAPTTKEPAPTTPK 439
QY 481 EPAPTAAPKPAATTPKEPAPTTPKEPAPTTTKESPTTPKEPAPTTYSAPTTKEPAT 540
DB 440 EPAPTAAPKPAATTPKEPAPTTPKEPAPTTTKESPTTPKEPAPTTYSAPTTKEPAT 499
QY 541 TTKSAPTTKESPTTPKEPAPTTPKEPAPTTPKKAPATTPKEPAPTTPKEPAPTTTK 600
DB 500 TTKSAPTTKESPTTPKEPAPTTPKEPAPTTPKKAPATTPKEPAPTTPKEPAPTTTK 559
QY 601 APTAKKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAT 660
DB 560 APTAKKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAT 619
QY 661 PEEPAATTPKAAAPNTPEKPAATTPKEPAATTPKEPAATTPKEPAATTPKEPATPTTLK 720
DB 620 PEEPAATTPKAAAPNTPEKPAATTPKEPAATTPKEPAATTPKEPAATTPKEPATPTTLK 679
QY 721 APTPKKAPKELAPTTKEBPTSTSDKAPPTTPKGTAPTTKEPAATTPKEPATPTTLK 780
DB 680 APTPKKAPKELAPTTKEBPTSTSDKAPPTTPKGTAPTTKEPAATTPKEPATPTTLK 739
QY 781 TAPTLKEPAATTPKPAKELAPTTKGTSTSDKAPPTTPKETAATTPKEPAATTPK 840
DB 740 TAPTLKEPAATTPKPAKELAPTTKGTSTSDKAPPTTPKETAATTPKEPAATTPK 799
QY 841 KPAPTPEPTPTTSEVSTPTTKKEPTTIHKSPEDESTBELSAPTPKALENSPKEGVPT 900
DB 800 KPAPTPEPTPTTSEVSTPTTKKEPTTIHKSPEDESTBELSAPTPKALENSPKEGVPT 859
QY 901 TKTPAATKPEMTTAKDITKEDRLATTPETTTAAPKMKETATTEKTESKITATTOV 960
DB 860 TKTPAATKPEMTTAKDITKEDRLATTPETTTAAPKMKETATTEKTESKITATTOV 919
QY 961 TSTTQDTPPKITTLKTTTLAPKVTYTKTITTEINMKKEETAKPKDRATNSKATTPK 1020
DB 920 TSTTQDTPPKITTLKTTTLAPKVTYTKTITTEINMKKEETAKPKDRATNSKATTPK 979
QY 1021 POKPTKAPKATSTYKPKPTMPRVKPKTTPTRKMTSTMPBELNPTSRIAEAMLQTTTBN 1080
DB 980 POKPTKAPKATSTYKPKPTMPRVKPKTTPTRKMTSTMPBELNPTSRIAEAMLQTTTBN 1039
QY 1081 QTPNSKLVENVPKSDAGAGETPHMLLRHVHPEKTPMDVYLRPNNGIITNPLMS 1140
DB 1040 QTPNSKLVENVPKSDAGAGETPHMLLRHVHPEKTPMDVYLRPNNGIITNPLMS 1099
QY 1141 DETNINCKGPDVGLITTLANGTLVAFRGHYFMMLSPSPSPSAPARRITTEVWGI 1200
DB 1100 DETNINCKGPDVGLITTLANGTLVAFRGHYFMMLSPSPSPSAPARRITTEVWGI 1159
QY 1201 RNCCEGKTFPFDKSOYWRFTNDIKDAGYKPKIFKFGGLTGOIYVAALSTANYKWPESVY 1260
DB 1160 RNCCEGKTFPFDKSOYWRFTNDIKDAGYKPKIFKFGGLTGOIYVAALSTANYKWPESVY 1219
QY 1261 FFKGGGSIQOYIYKOEPRQKGRRRPALNVYGEEMOVRARRRPRALGSPQTHIRIOY 1320
DB 1220 FFKGGGSIQOYIYKOEPRQKGRRRPALNVYGEEMOVRARRRPRALGSPQTHIRIOY 1279
QY 1321 SPARLAYODKGVLANEVKSLIMRGLPNVYTSALSLPIRRPDGDYVAFKQOYNNIDV 1380
DB 1280 SPARLAYODKGVLANEVKSLIMRGLPNVYTSALSLPIRRPDGDYVAFKQOYNNIDV 1339
QY 1381 PSRTARAITTSQOTLSKVMYNCP 1404
DB 1340 PSRTARAITTSQOTLSKVMYNCP 1363

RESULT 5
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine

```


RESULT 6
US-10-124-557-46
Sequence 46, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Hewick, Rodney M.
Gessner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cbeert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 93.2%; Score 7011; DB 14; Length 1320;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAMKTIPIYLLLSVFIQVSSODLSACGRGEGYRDATCNCDYNCQHYMECCPDP 60
DB 1 MAMKTIPIYLLLSVFIQVSSO----- 25
QY 61 KRVCTALSLCKGKCFPSFERGRCDDAOCKKTKDKCCPDYSEFCALVHNFTSPSSKAP 120
DB 26 -----ELSCGKRCFSPFERGRCDDAOCKKTKDKCCPDYSEFCALVHNFTSPSSKAP 79
QY 121 PPSGAGOTTKSTTKRSPKPNKKTKKVLSESEITEHSVSENGSSSSSSSSSSSTIM 180
DB 80 PPSGAGOTTKSTTKRSPKPNKKTKKVLSESEITE----- 115
QY 181 KIKSKIKSAMRELQKKLVKDNKKKRTKKKPTPKPVVDGAGSLDNGDFKVTTPDTST 240
DB 115 KIKSKIKSAMRELQKKLVKDNKKKRTKKKPTPKPVVDGAGSLDNGDFKVTTPDTST 240

DB 116 -----VKDNKKRRTKKKPTPKPVVDGAGSLDNGDFKVTTPDTST 156
QY 241 TQHNKYSTPKITTAAPINRPSLPNSDTSKETSITVKKETVETKETTNNKQSTOG 300
DB 157 TQHNKYSTPKITTAAPINRPSLPNSDTSKETSITVKKETVETKETTNNKQSTOG 216
QY 301 KEKTSKETSQSIKTSKADLAFTSKVLAKFPKAEITTKGPAITTKPEPTPTTKPEAS 360
DB 217 KEKTSKETSQSIKTSKADLAFTSKVLAKFPKAEITTKGPAITTKPEPTPTTKPEAS 216
QY 361 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
DB 277 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 336
QY 421 APTTTSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 480
DB 337 APTTTSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 396
QY 481 EPAPTTAPKKAPATTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 540
DB 397 EPAPTTAPKKAPATTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 456
QY 541 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 600
DB 457 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 516
QY 601 APTAPKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAP 660
DB 517 APTAPKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAP 576
QY 661 PEEBAPTTKKAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 720
DB 577 PEEBAPTTKKAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 636
QY 721 APTPKKPAKELAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 780
DB 637 APTPKKPAKELAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 696
QY 781 TAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 840
DB 697 TAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 756
QY 841 KAPPTPEPTPTTSVSTPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 900
DB 757 KAPPTPEPTPTTSVSTPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 816
QY 901 TKTPATKPEMTTAKDKTERDLRTTPTTAAAPMTKETAATTEKTESKITATTTQV 960
DB 817 TKTPATKPEMTTAKDKTERDLRTTPTTAAAPMTKETAATTEKTESKITATTTQV 876
QY 961 TSTTTODITTPKITTLLKTTTLPKTTTKKITTTEIANKPBEITAKKORATNSKATTPK 1020
DB 877 TSTTTODITTPKITTLLKTTTLPKTTTKKITTTEIANKPBEITAKKORATNSKATTPK 936
QY 1021 POKPTKAPKKPTSTKKPKTPRVKPKTTPTRKMTSTMPBLNPTSRIEAMQTTTRN 1080
DB 937 POKPTKAPKKPTSTKKPKTPRVKPKTTPTRKMTSTMPBLNPTSRIEAMQTTTRN 996
QY 1081 QTPNSKLVEVNRKSEBAGABETPMLLRPHVMEVTPPDMDYLPRVNOGIIINPMS 1140
DB 997 QTPNSKLVEVNRKSEBAGABETPMLLRPHVMEVTPPDMDYLPRVNOGIIINPMS 1056
QY 1141 DETNINCGRVUGLTLNNGTILVATRGHYFMMLSPSPSPSPARRITTEWGISPIDVT 1200
DB 1057 DETNINCGRVUGLTLNNGTILVATRGHYFMMLSPSPSPSPARRITTEWGISPIDVT 1116
QY 1201 RCNCEGKTFEFDOSQVRFNTNDIKDAGVKKPIFKGGGGLTGQIVAAALSTAKYKMMESY 1260
DB 1117 RCNCEGKTFEFDOSQVRFNTNDIKDAGVKKPIFKGGGGLTGQIVAAALSTAKYKMMESY 1176
QY 1261 PFKRGSIIQVYIKQEPVQKCGRRPALNYPVYGEWTVQRRRRFERAIGSQTHTRIOY 1320
DB 1177 PFKRGSIIQVYIKQEPVQKCGRRPALNYPVYGEWTVQRRRRFERAIGSQTHTRIOY 1236

QY 1321 SPARLAYODKGVLEHNEVKVSIIMRGLPNVVTSAISLPMIRKPDGYDYAFAKDOYVNDV 1380
DB 1237 SPARLAYODKGVLEHNEVKVSIIMRGLPNVVTSAISLPMIRKPDGYDYAFAKDOYVNDV 1296
QY 1381 PSRTARAITTSGQTLISKVWYNCP 1404
DB 1297 PSRTARAITTSGQTLISKVWYNCP 1320

RESULT 7
US-10-124-557-60
Sequence 60, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124.557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caer, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match 93.2%; Score 7011; DB 14; Length 1320;
Best Local Similarity 94.0%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAMWTLPIYLLLSVFVIOVSSODLSSCAGRCGEGYGRDATCNCDVNCQHWECDF 60
DB 1 MAMWTLPIYLLLSVFVIOVSSODLSSCAGRCGEGYGRDATCNCDVNCQHWECDF 25
QY 61 KRVCTAELSCGRGCFESFERGECDCDAQCCKYDKCCPDYESFCAEVNPTSPSSKKAP 120

DB 26 -----ELSCGRGCFESFERGECDCDAQCCKYDKCCPDYESFCAEVNPTSPSSKKAP 79
QY 121 PPSGASQITKSTTKRSRPPPKKTKKYIESEETIEHSVENQSSSSSSSSSTW 180
DB 80 PPSGASQITKSTTKRSRPPPKKTKKYIESEETIEHSVENQSSSSSSSSSTW 115
QY 181 KIKSSKNSANRELOKLLKVNQNKKNRTKKKPTPKRPVVDAGSGLDJNGDFKVTTPDST 240
DB 116 -----VDNKNKRTKKKPTPKRPVVDAGSGLDJNGDFKVTTPDST 156
QY 241 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSILVKNKETEYVETKTTTTNKTSTDG 300
DB 157 TQHNKVSPTKITTAKPINRPSLPNSDTSKETSILVKNKETEYVETKTTTTNKTSTDG 216
QY 301 KEKTSKRNOSIKRTSKOLAPTSKYLAKTTPKAEITTKPALTTKPEPTTPPKBAS 360
DB 217 KEKTSKRNOSIKRTSKOLAPTSKYLAKTTPKAEITTKPALTTKPEPTTPPKBAS 276
QY 361 TTPKEPTTPPKBASPTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAP 420
DB 277 TTPKEPTTPPKBASPTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAP 336
QY 421 APITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 480
DB 337 APITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 396
QY 481 EPAPTAAPKAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 540
DB 397 EPAPTAAPKAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 456
QY 541 TTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 600
DB 457 TTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 516
QY 601 APAPKAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 660
DB 517 APAPKAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 576
QY 661 PEPAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 720
DB 577 PEPAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 636
QY 721 APITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 780
DB 637 APITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 696
QY 781 TAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 840
DB 697 TAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 756
QY 841 KPAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 900
DB 757 KPAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 816
QY 901 TKTAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 960
DB 817 TKTAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAPITTKSAP 876
QY 961 TSTTQDTTPPKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTK 1020
DB 877 TSTTQDTTPPKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTKITTK 936
QY 1021 POKPTAPKPKPTSKKKTTPKPKPTSKKKTTPKPKPTSKKKTTPKPKPTSKKKTTPKPKPTSKKKTTPK 1080
DB 937 POKPTAPKPKPTSKKKTTPKPKPTSKKKTTPKPKPTSKKKTTPKPKPTSKKKTTPKPKPTSKKKTTPK 996
QY 1081 QTPNSKLVENPKSEDEGAGETPHMLLRPHVMPVETPDMDLPRVPMOGLIINPMLS 1140
DB 997 QTPNSKLVENPKSEDEGAGETPHMLLRPHVMPVETPDMDLPRVPMOGLIINPMLS 1056
QY 1141 DETNICKGKPVDTLTLRNGTLVAFRGHYFMMLSPSPSPSAPRITTEWGIISPDIQVFT 1200
DB 1057 DETNICKGKPVDTLTLRNGTLVAFRGHYFMMLSPSPSPSAPRITTEWGIISPDIQVFT 1116

QY 1201 RCNCEGTFFPKDSOYWRFTNDIKDAGYPRKIFKFGGLTGOIYVAAALSTAKYKNWPSVY 1260
DB 1117 RCNCEGTFFPKDSOYWRFTNDIKDAGYPRKIFKFGGLTGOIYVAAALSTAKYKNWPSVY 1176
QY 1261 FFRGGSIOQYIYKOEBOVKCPGRPALNYPVGEOTVARRRPERAIGPSQHTIRIQY 1320
DB 1177 FFRGGSIOQYIYKOEBOVKCPGRPALNYPVGEOTVARRRPERAIGPSQHTIRIQY 1236
QY 1321 SPRLAYQDKGVLANEYKVSILWRGLPNVTSALSPNIRKPDGYDYAASKOQYNNIDY 1380
DB 1237 SPRLAYQDKGVLANEYKVSILWRGLPNVTSALSPNIRKPDGYDYAASKOQYNNIDY 1296
QY 1381 PSRTARAITTRSGOTLSKWMYNC 1404
DB 1297 PSRTARAITTRSGOTLSKWMYNC 1320

RESULT 8
US-10-124-557-42
Sequence 42, Application US/10124557
Publication NO. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gaetz, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match 93.4%; Score 7007.5; DB 14; Length 1311;
Best Local Similarity 93.4%; Pred. No. 0;

Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAMKTLPIYLLILLISFVYIQOVSQDLSACAGCGGYSRDATCNDYNCQHYMECCPDP 60
DB 1 MAMKTLPIYLLILLISFVYIQOVSQDLSACAGCGGYSRDATCNDYNCQHYMECCPDP 60
QY 61 KKVCTAELSCGRCPFSFGRGRCDDAQCCKYDKCCPYESFCAEVHNPTSPSSKAP 120
DB 61 KKVCTAELSCGRCPFSFGRGRCDDAQCCKYDKCCPYESFCAEVHNPTSPSSKAP 106
QY 121 PPSGASQIRIKSTTKSPKPPNKKTKVIESEITIEHSVSENOESSSSSSSSSTIWM 180
DB 107 ----- 106
QY 181 KIKSSKNSAANBELQKLVKONKKNRTKKKTPKPPVVDAGSGIDNDGFKYTPDTS 240
DB 107 -----VKONKKNRTKKKTPKPPVVDAGSGIDNDGFKYTPDTS 147
QY 241 TOHNKVSTSPKITTAKPINRPSLPPNSDTSKETSILVNKEITVETKETTNNKOTSDG 300
DB 148 TOHNKVSTSPKITTAKPINRPSLPPNSDTSKETSILVNKEITVETKETTNNKOTSDG 207
QY 301 KKKTSASAKTOSIEKTSANDLAPTSKVLAKEPTTKGPAITTPKKEPTTPPKEPAS 360
DB 208 KKKTSASAKTOSIEKTSANDLAPTSKVLAKEPTTKGPAITTPKKEPTTPPKEPAS 267
QY 361 TTPKEPTPTTISAPITPKKPAITTPKSAITTPKKEPAITTPKKEPAITTPKKEP 420
DB 268 TTPKEPTPTTISAPITPKKPAITTPKSAITTPKKEPAITTPKKEPAITTPKKEP 327
QY 421 APTTKSAITTPKKEPAITTPKKAAPITTPKKEPAITTPKKEPTTPPKKEPAITTPK 480
DB 328 APTTKSAITTPKKEPAITTPKKAAPITTPKKEPAITTPKKEPTTPPKKEPAITTPK 387
QY 481 BPAPTAAPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 540
DB 388 BPAPTAAPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 447
QY 541 TTKSAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 600
DB 448 TTKSAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 507
QY 601 APTAPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 660
DB 508 APTAPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 567
QY 661 PEEPAITTPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 720
DB 568 PEEPAITTPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 627
QY 721 APTTPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 780
DB 628 APTTPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 687
QY 781 TAPITLKEPAITTPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 840
DB 688 TAPITLKEPAITTPKKAAPITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 747
QY 841 KPAITTPETPTTSEVSTPTTKEPTTIHKSPPDSSTPELSAAPTPKALENSPKERGVT 900
DB 748 KPAITTPETPTTSEVSTPTTKEPTTIHKSPPDSSTPELSAAPTPKALENSPKERGVT 807
QY 901 TITPAATKPEMTTAKDKTTERDLRTTBETTTAAKMTKSTIYTTKEPTTESKITATTV 960
DB 808 TITPAATKPEMTTAKDKTTERDLRTTBETTTAAKMTKSTIYTTKEPTTESKITATTV 867
QY 961 TSTTQDTPPKITTLTKTTTAPKVYTTTKKITTEINMKPESTAKXORANSTASATPK 1020
DB 868 TSTTQDTPPKITTLTKTTTAPKVYTTTKKITTEINMKPESTAKXORANSTASATPK 927
QY 1021 POKPTKAPKKPTSTKPKTMRVRKPTTPPRKMTSTMBELNPSRIAEMALQTTTREN 1080
DB 928 POKPTKAPKKPTSTKPKTMRVRKPTTPPRKMTSTMBELNPSRIAEMALQTTTREN 987

QY 1081 QTPNSKLYEVNPKSDAGAGSETPHMLLRPHVEMPEVTPMDYLPRVNOGIIINPMLS 1140
DB 988 QTPNSKLYEVNPKSDAGAGSETPHMLLRPHVEMPEVTPMDYLPRVNOGIIINPMLS 1047
QY 1141 DETNINCKGKPVDPGLTTLNGLTLVAFRGHYFMWLSPPSPSPARRITTEVWGIIPSPIDVFT 1200
DB 1048 DETNINCKGKPVDPGLTTLNGLTLVAFRGHYFMWLSPPSPSPARRITTEVWGIIPSPIDVFT 1107
QY 1201 RCNCEGKTPFFKDSQYMFNFNDIKDAGYPKPIFKGFGGLTGQIVAASTAKYKMPESVY 1260
DB 1108 RCNCEGKTPFFKDSQYMFNFNDIKDAGYPKPIFKGFGGLTGQIVAASTAKYKMPESVY 1167
QY 1261 FFKGGSIIQYIYKQEPVQKCGRRPALNYPYGEVTOVRRRRRPERAIGPSQTHIRIQY 1320
DB 1168 FFKGGSIIQYIYKQEPVQKCGRRPALNYPYGEVTOVRRRRRPERAIGPSQTHIRIQY 1227
QY 1321 SPARLAYODKVLHNEVKVSLIMRGLPNVVTSAISLPNIRKPDGYDYAFSNDQYNNIDV 1380
DB 1228 SPARLAYODKVLHNEVKVSLIMRGLPNVVTSAISLPNIRKPDGYDYAFSNDQYNNIDV 1287
QY 1381 PSRTARAITTRSGQTLGKVMYNCP 1404
DB 1287 PSRTARAITTRSGQTLGKVMYNCP 1311

RESULT 9
US-10-124-557-50
Sequence 50, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124.557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseer, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50
Query Match 92.6%; Score 6971; DB 14; Length 1314;
Best Local Similarity 93.6%; Pred. No. 0;
Matches 1314; Conservative 0; Mismatches 0; Indels 90; Gaps 1;
QY 1 MAMKTLPIYLLLLSVPIQQVSSQDLSGCGRGEYSRDATCNDYNCQHMECCDF 60
DB 1 MAMKTLPIYLLLLSVPIQQVSSQDLSGCGRGEYSRDATCNDYNCQHMECCDF 60
QY 61 KRVCIAELSCGRCFESFERREDCDAQCKKVKCCPDYSPCAEVNPFSPSSKAP 120
DB 61 KRVCIAELSCGRCFESFERREDCDAQCKKVKCCPDYSPCAEVNPFSPSSKAP 66
QY 121 PPSGASQTIKSTYKSPKPKKKTKVIESEIIEHSVENQESSSSSSSSSTIWM 180
DB 67 -----EHSVENQESSSSSSSSSTIWM 90
QY 181 KIKSSKNSAANRELQKLVKDNKNKRTKKKPTPKPVVDAGSGLDNGDFKVTPTDST 240
DB 91 KIKSSKNSAANRELQKLVKDNKNKRTKKKPTPKPVVDAGSGLDNGDFKVTPTDST 150
QY 241 TQHNKYSTSPKITAKINRPRLPNSDTSKETSIVNKKETVETKETTNNKQSTDG 300
DB 151 TQHNKYSTSPKITAKINRPRLPNSDTSKETSIVNKKETVETKETTNNKQSTDG 210
QY 301 KEKTTSAKETOSIKTSAKDLAPTSKYLAKPTPAETTTGAPALTTPKEPTTPKEPAS 360
DB 211 KEKTTSAKETOSIKTSAKDLAPTSKYLAKPTPAETTTGAPALTTPKEPTTPKEPAS 270
QY 361 TTPKEPTTPKSAPTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 420
DB 271 TTPKEPTTPKSAPTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 330
QY 421 APTTTKAPPTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 480
DB 331 APTTTKAPPTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK 390
QY 481 EPAPTAKKKAPPTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 540
DB 391 EPAPTAKKKAPPTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 450
QY 541 TTKSAPPTPKESPPTTKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 600
DB 451 TTKSAPPTPKESPPTTKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 510
QY 601 APTAPKBPAPPTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTP 660
DB 511 APTAPKBPAPPTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTP 570
QY 661 PEEBAPTTPKAAANTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 720
DB 571 PEEBAPTTPKAAANTPKBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPT 630
QY 721 APTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGAAPTTPKKEBAPTTPKKEBAPT 780
DB 631 APTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKGAAPTTPKKEBAPTTPKKEBAPT 690
QY 781 TAPTTLKEBAPTTPKKAPELAPTTTKEPTSTSDKAPATTTPKGAAPTTPKKEBAPTTPK 840
DB 691 TAPTTLKEBAPTTPKKAPELAPTTTKEPTSTSDKAPATTTPKGAAPTTPKKEBAPTTPK 750
QY 841 KPAPTTETPTPTTSEVSTPTTKEPTTIKSPDSETPELASBPTPALENSPEKPVPT 900
DB 751 KPAPTTETPTPTTSEVSTPTTKEPTTIKSPDSETPELASBPTPALENSPEKPVPT 810
QY 901 TKTPPAATKPEWTTTAKDXTTERDLRTTETTTAAPKATKETAATTEKTTESKITAATTV 960
DB 811 TKTPPAATKPEWTTTAKDXTTERDLRTTETTTAAPKATKETAATTEKTTESKITAATTV 870

QY 961 TSTTODTTPKRTTTLTKTTLAKRTTTLTKTTLTTEIMNKPEETAKPKRATNSKATTPK 1020
 DB 871 TSTTODTTPKRTTTLTKTTLAKRTTTLTKTTLTTEIMNKPEETAKPKRATNSKATTPK 930
 QY 1021 POKPTAPKPKPTSTKPKPTMPRVKPKPTTTPPKMTSTMBELNPTSRIAEAMLQTTTRN 1080
 DB 931 POKPTAPKPKPTSTKPKPTMPRVKPKPTTTPPKMTSTMBELNPTSRIAEAMLQTTTRN 990
 QY 1081 QTPNSKLVEVNPKEBDAAGAEGETPHMLRPHFMPEBVTTPDMYLRVRPQGIINPMIS 1140
 DB 991 QTPNSKLVEVNPKEBDAAGAEGETPHMLRPHFMPEBVTTPDMYLRVRPQGIINPMIS 1050
 QY 1141 DETNINCKRPVDGLTTLRNGTLVAFRGHYFMMLSPSPSPAPARITREWGIPSPIDTVP 1200
 DB 1051 DETNINCKRPVDGLTTLRNGTLVAFRGHYFMMLSPSPSPAPARITREWGIPSPIDTVP 1110
 QY 1201 RCNCEGKTFEPKDSQYWRFTNDIKDAGYPRKPIKGFGLTGQIVAAALSTAKYKMBESY 1260
 DB 1111 RCNCEGKTFEPKDSQYWRFTNDIKDAGYPRKPIKGFGLTGQIVAAALSTAKYKMBESY 1170
 QY 1261 PFRGGSIOQYIYKQEPVQKCPGRPALNPTVYGEHTQVRRRFBPAIGPSQTHTRIOY 1320
 DB 1171 PFRGGSIOQYIYKQEPVQKCPGRPALNPTVYGEHTQVRRRFBPAIGPSQTHTRIOY 1230
 QY 1321 SPARLAYODKGVLANEVKVSILWRGLPNVYTSALSPNIRKPDGYDYAFSKOQYNNIDV 1380
 DB 1231 SPARLAYODKGVLANEVKVSILWRGLPNVYTSALSPNIRKPDGYDYAFSKOQYNNIDV 1290
 QY 1381 PSRTARAITTRSGQTLKSWYNNCP 1404
 DB 1291 PSRTARAITTRSGQTLKSWYNNCP 1314

RESULT 10
 US-10-124-557-142
 Sequence 142, Application US/10124557
 Publication No. US20020137894A1
 GENERAL INFORMATION:
 APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Geener, Thomas G.
 TITLE OR INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetice Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cbeert, Luann

REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 142:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1313 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 142:
 US-10-124-557-142
 Query Match 92.4%; Score 6951.5; DB 14; Length 1313;
 Best Local Similarity 93.5%; Pred. No. 0;
 Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
 1 MAMKTLPIYLILLLVFVLIQVSSODLSSCAGRCGEYSRDATCNDYNCQHMECCPDF 60
 1 MAMKTLPIYLILLLVFVLIQVSSODLSSCAGRCGEYSRDATCNDYNCQHMECCPDF 25
 61 KKVCTAELSCRCESFERGECDDAQCKKYDKCCPDYSFCAEVHNPTSPSSKAP 120
 26 -----ELSCRCESFERGECDDAQCKKYDKCCPDYSFCA----- 64
 121 PPSGASQITTKTSKPRPKKTKVIESEITEHSVSNQSSSSSSSSSSSSSTTW 180
 65 -----EHSVSNQSSSSSSSSSSSSSSSTTW 89
 181 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPVVDAGSGLDNDGFVYTPPDST 240
 90 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPVVDAGSGLDNDGFVYTPPDST 149
 241 TQHNKVSSTPKITTKAPINRPSLPNSDTSKETSIVNKEITVEKETTNNKQSTIDG 300
 150 TQHNKVSSTPKITTKAPINRPSLPNSDTSKETSIVNKEITVEKETTNNKQSTIDG 209
 301 KEKTTSAKGTOSIEKTSKDLAPTSKVLAKPTKAKETTTKGALTTPKSEPTTTKEPAS 360
 210 KEKTTSAKGTOSIEKTSKDLAPTSKVLAKPTKAKETTTKGALTTPKSEPTTTKEPAS 269
 361 TTPKEPTPTTKISAPTTPEKAPATTTPKSAPTTPKBPATTTPKBPATTTPKBP 420
 270 TTPKEPTPTTKISAPTTPEKAPATTTPKSAPTTPKBPATTTPKBPATTTPKBP 329
 421 APTTKSAPTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBP 480
 330 APTTKSAPTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBP 389
 481 BPAPTAAPKPAATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATT 540
 390 BPAPTAAPKPAATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATT 449
 541 TTKSAPTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATT 600
 450 TTKSAPTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATT 509
 601 APTAKBPAPTTPKETAATTPKGLNPTPEKLAATTTPKBPATTTPKBPATTTPKBP 660
 510 APTAKBPAPTTPKETAATTPKGLNPTPEKLAATTTPKBPATTTPKBPATTTPKBP 569
 661 PEEPAPTPKAAAPNTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATT 720
 570 PEEPAPTPKAAAPNTPKBPATTTPKBPATTTPKBPATTTPKBPATTTPKBPATT 629
 721 APTTKKAPKELAPTTTKEPTSTSDKAPATTTPKGAATTTPKBPATTTPKBPATT 780
 630 APTTKKAPKELAPTTTKEPTSTSDKAPATTTPKGAATTTPKBPATTTPKBPATT 689
 781 TAPTLKEBAPTTPKKAPKELAPTTTGTGPTSTSDKAPATTTPKETAATTPKBPATT 840
 690 TAPTLKEBAPTTPKKAPKELAPTTTGTGPTSTSDKAPATTTPKETAATTPKBPATT 749

QY 841 KPAATPEPTSEVSTPTTKKPTTIHKSPODESTBELSAEPKALENSPEKPGPT 900
DB 750 KPAATPEPTSEVSTPTTKKPTTIHKSPODESTBELSAEPKALENSPEKPGPT 809
QY 901 TKTPEATPEMTTAKOKTEREDLRTTPELTAAAPKMTKETATTTTEKTESKITATTTQV 960
DB 810 TKTPEATPEMTTAKOKTEREDLRTTPELTAAAPKMTKETATTTTEKTESKITATTTQV 869
QY 961 TSTTTODTTPPKITTLKTTLTTLAPKVTYTKITITTEINAKKBEETAKPDRATNSKATTPK 1020
DB 870 TSTTTODTTPPKITTLKTTLTTLAPKVTYTKITITTEINAKKBEETAKPDRATNSKATTPK 929
QY 1021 POKTKAPKPTSTTKPKTPMVRKPKTPTPKMTSTMPLELNPSTRIAEAMLOTTREPN 1080
DB 930 POKTKAPKPTSTTKPKTPMVRKPKTPTPKMTSTMPLELNPSTRIAEAMLOTTREPN 989
QY 1081 QTPNSKIVEVNPKESEDAGAGETPHMLLRPHVEMPEVTPTMDYLLPRVNPQGIINPMLS 1140
DB 990 QTPNSKIVEVNPKESEDAGAGETPHMLLRPHVEMPEVTPTMDYLLPRVNPQGIINPMLS 1049
QY 1141 DETNINCKGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPSPARRITTEVWGIIPSPIDVTFT 1200
DB 1050 DETNINCKGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPSPARRITTEVWGIIPSPIDVTFT 1109
QY 1201 RCNCEGKTFPPKDSQYMRFTNDIDAGYPKPIFKSGFGLTGQIYVAALSTAKYKMMPESVY 1260
DB 1110 RCNCEGKTFPPKDSQYMRFTNDIDAGYPKPIFKSGFGLTGQIYVAALSTAKYKMMPESVY 1169
QY 1261 FFKGSGSIQOYIYKQEPVQKCPGRPALNVPYGEVMTQVRRRRFERAIGPSQTHIRIQY 1320
DB 1170 FFKGSGSIQOYIYKQEPVQKCPGRPALNVPYGEVMTQVRRRRFERAIGPSQTHIRIQY 1229
QY 1321 SPARLAYODKVLNNEVAVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKOYNNIDV 1380
DB 1230 SPARLAYODKVLNNEVAVSILMRGLPNVVTSAISLPIRKPDGYDYAFSKOYNNIDV 1289
QY 1381 PSRTARATITRSQGTLSKVMTNCP 1404
DB 1290 PSRTARATITRSQGTLSKVMTNCP 1313

RESULT 11
US-10-124-557-44
Sequence 44, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44
Query Match 89.4%; Score 6727; DB 14; Length 1270;
Best Local Similarity: 90.5%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
QY 1 MAMKTLPIYLLLSVIVIOVSSODLSSCAGRGEGYSRATCNQCHVMECCPDF 60
DB 1 MAMKTLPIYLLLSVIVIOVSSQ----- 25
QY 61 KRVCIAELSCKGRCFESFERGECDCDQCKYDKCCPDYESFCAEVNPTSPSSKAP 120
DB 26 -----ELSCKGRCFESFERGECDCDQCKYDKCCPDYESFCAE----- 65
QY 121 PPGASQITKSTTRSPKPKKKYKVIIESEITIEHSVSENGESSSSSSSSSTIW 180
DB 66 ----- 65
QY 181 KIKSKNSAANRELOKLLKYNKKNKTKKKPTPKPVVNDAGSGLDNGDPKVTTPPTST 240
DB 66 -----VDNKNKNTKKKPTPKPVVNDAGSGLDNGDPKVTTPPTST 106
QY 241 TQHNKSTSPKLTAKINPRLPSLPNSDTSKETSLTVNKEITYETKETTTNKTQSTDG 300
DB 107 TQHNKSTSPKLTAKINPRLPSLPNSDTSKETSLTVNKEITYETKETTTNKTQSTDG 166
QY 301 KEKTTSAKETQSIKETSADOLAPTSKVLAKPTPAETTTKGPALTTPKEPTTPKEPAS 360
DB 167 KEKTTSAKETQSIKETSADOLAPTSKVLAKPTPAETTTKGPALTTPKEPTTPKEPAS 226
QY 361 TTPKEPTTPKSAPTTPKEPAFTTTSAPPTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 420
DB 227 TTPKEPTTPKSAPTTPKEPAFTTTSAPPTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 286
QY 421 APTTTSAPPTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 480
DB 287 APTTTSAPPTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 346
QY 481 EPAFTAPKPAFTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 540
DB 347 EPAFTAPKPAFTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 406
QY 541 TTKSAPPTPKPSPTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 600
DB 407 TTKSAPPTPKPSPTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 466
QY 601 APTAPKPAFTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 660
DB 467 APTAPKPAFTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 526
QY 661 PEEBAPPTPKAAANTPKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 720
DB 527 PEEBAPPTPKAAANTPKPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKEPAFTTTPKE 586

QY 721 APTTPKPAKELAPTTTKEPTSTSDKPAATTPKGAATTPKPAATTPKKEPAPTTPKG 780
DB 587 APTTPKPAKELAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPKG 646
QY 781 TAPTTKEPAPTTKPAKELAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPK 840
DB 647 TAPTTKEPAPTTKPAKELAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPK 706
QY 841 KPAATTPKPAATTPKKEPAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPK 900
DB 707 KPAATTPKPAATTPKKEPAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPK 766
QY 901 TAPTTKEPAPTTTAKDRTTAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPK 960
DB 767 TAPTTKEPAPTTTAKDRTTAPTTTKEPTSTSDKPAATTPKGAATTPKKEPAPTTPK 826
QY 961 TSTTTODTTPPKITTLKTTTTLAPKVTYTKKTTTTEIMNKPEETAKPKDRTNSKATTPK 1020
DB 827 TSTTTODTTPPKITTLKTTTTLAPKVTYTKKTTTTEIMNKPEETAKPKDRTNSKATTPK 886
QY 1021 POKPTAPKPKPTSTKPKTPKTPKPKTPKTPKTPKTPKTPKTPKTPKTPKTPKTPKTPK 1080
DB 887 POKPTAPKPKPTSTKPKTPKTPKPKTPKTPKTPKTPKTPKTPKTPKTPKTPKTPKTPK 946
QY 1081 QTPNSKLVEVNPSEBAGAGAGETPHMLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140
DB 947 QTPNSKLVEVNPSEBAGAGAGETPHMLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006
QY 1141 DETNINCKRPVDTLTLRNGTLVAFRGHYFMMLSPSPSPARITTEWGISPIDTVP 1200
DB 1007 DETNINCKRPVDTLTLRNGTLVAFRGHYFMMLSPSPSPARITTEWGISPIDTVP 1066
QY 1201 RNCCEKTEPPKQSOQWRFTNDIKDAGYKPKIFKGGGLTGOVALSTAKTKNMBESY 1260
DB 1067 RNCCEKTEPPKQSOQWRFTNDIKDAGYKPKIFKGGGLTGOVALSTAKTKNMBESY 1126
QY 1261 PFKRGGSIQOYIYKQBPVOKCPGRPALNYPVGEVTOYRRRFEALIGPSQHTIRIY 1320
DB 1127 PFKRGGSIQOYIYKQBPVOKCPGRPALNYPVGEVTOYRRRFEALIGPSQHTIRIY 1186
QY 1321 SPARLAVODKGVLANEVKVSILMRGLPNVNTSAISLPIRKPDGYDYVAFSKQOYINIV 1380
DB 1187 SPARLAVODKGVLANEVKVSILMRGLPNVNTSAISLPIRKPDGYDYVAFSKQOYINIV 1246
QY 1381 PSRTAAITTRSGQITLSKWMYNC 1404
DB 1247 PSRTAAITTRSGQITLSKWMYNC 1270

RESULT 12
US-10-124-557-104
Sequence 104, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clerk, Stephen C.
Jacob, Kenneth
Hewick, Rodney M.
Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104
Query Match 80.9%; Score 6085; DB 14; Length 1140;
Best Local Similarity 100.0%; Pred. 5.8e-302; Indels 0; Gaps 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NAMKTLPIYLLLSVFYIQVSSQDLSACAGCEGYSRDATCNDYNCQHYMECCPDF 60
DB 1 NAMKTLPIYLLLSVFYIQVSSQDLSACAGCEGYSRDATCNDYNCQHYMECCPDF 60
QY 61 KAVCTAELSCKRCESFERGECDDAQCKKYDCCPDYEFCAEVNPTSPSSKAP 120
DB 61 KAVCTAELSCKRCESFERGECDDAQCKKYDCCPDYEFCAEVNPTSPSSKAP 120
QY 121 PPSGASQITIKSTTKSPKPPNKKTKYIESBEITEBHSVSENQSSSSSSSSSTIM 180
DB 121 PPSGASQITIKSTTKSPKPPNKKTKYIESBEITEBHSVSENQSSSSSSSSSTIM 180
QY 181 KIKSSKNSAANRELQKLVKXONKXRTKKPTPKRPVVDAGSGLDNGDFKVTTPDTST 240
DB 181 KIKSSKNSAANRELQKLVKXONKXRTKKPTPKRPVVDAGSGLDNGDFKVTTPDTST 240
QY 241 TQHNKVSISPKITTKAPINPRPSLPNSDTSKETSITVANKETTVETKETTNNKOTSDG 300
DB 241 TQHNKVSISPKITTKAPINPRPSLPNSDTSKETSITVANKETTVETKETTNNKOTSDG 300
QY 301 KEKTTSAKBTOSIEKTSKADLAPTSKYLAKPPKATTTKGPALTTPKEPTTPKBPAS 360
DB 301 KEKTTSAKBTOSIEKTSKADLAPTSKYLAKPPKATTTKGPALTTPKEPTTPKBPAS 360
QY 361 TTPKEPTTPKSAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
DB 361 TTPKEPTTPKSAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY 421 APTTKSAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480
DB 421 APTTKSAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480
QY 481 BPAPTPAPKPAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540
DB 481 BPAPTPAPKPAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540
QY 541 TTKSAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600
DB 541 TTKSAATTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600

QY	601	APLAPKEBAPPTPKETAJPTPKKLLPPTTPEKALPPTPEKAPPTPEBELAPPTPEBPTPTT	660
Db	601	APLAPKEBAPPTPKETAJPTPKKLLPPTTPEKALPPTPEKAPPTPEBELAPPTPEBPTPTT	660
QY	601	APLAPKEBAPPTPKETAJPTPKKLLPPTTPEKALPPTPEKAPPTPEBELAPPTPEBPTPTT	660
QY	661	PEBPAPPTPKAAAPPTPEKBPAPPTTKEBPAPPTPKBPAPPTPKETAJPTTPKGTAPPTTKEP	720
Db	661	PEBPAPPTPKAAAPPTPEKBPAPPTTKEBPAPPTPKBPAPPTPKETAJPTTPKGTAPPTTKEP	720
QY	721	APPTPEKBPAPKEIAPPTTKEPPTSTSDBPAPPTPKGTAPPTPEKBPAPPTPEKBPAPPTPEK	780
Db	721	APPTPEKBPAPKEIAPPTTKEPPTSTSDBPAPPTPKGTAPPTPEKBPAPPTPEKBPAPPTPEK	780
QY	781	TAPPTTKEBPAPPTPKKBPAPKEIAPPTTGTPTSTSDKBPAPPTPKETAJPTPEKBPAPPTPK	840
Db	781	TAPPTTKEBPAPPTPKKBPAPKEIAPPTTGTPTSTSDKBPAPPTPKETAJPTPEKBPAPPTPK	840
QY	841	KPAPPTPEBPPTTSEVSTPTTKEPPTTIHKSPTDSBELSAPPKKLSPKSGVPT	900
Db	841	KPAPPTPEBPPTTSEVSTPTTKEPPTTIHKSPTDSBELSAPPKKLSPKSGVPT	900
QY	901	TKTPAPATPEMTTAKDITTEBDLTTTBETTTAABKMTKETATTTTEKTESKITATTTQV	960
Db	901	TKTPAPATPEMTTAKDITTEBDLTTTBETTTAABKMTKETATTTTEKTESKITATTTQV	960
QY	961	TSTTTQDPTPEKITLKTITLLAPKTTTKKITTTTEINNKPEPTAKKBPATNSAATPK	1020
Db	961	TSTTTQDPTPEKITLKTITLLAPKTTTKKITTTTEINNKPEPTAKKBPORATNSAATPK	1020
QY	1021	POKPTPAKPKPTSTCKPKTPBRVRKPTTTPBRKMTSTMPBELNPTSRITAEAMLTQTTTBN	1080
Db	1021	POKPTPAKPKPTSTCKPKTPBRVRKPTTTPBRKMTSTMPBELNPTSRITAEAMLTQTTTBN	1080
QY	1081	QTPNSKLYENVKPSDAGABEETHMLLRPHVEMBPATPMDDYAPRVNCGIILNPLMS	1140
Db	1081	QTPNSKLYENVKPSDAGABEETHMLLRPHVEMBPATPMDDYAPRVNCGIILNPLMS	1140

RESULT 13-557-58
US-10-124-557-58
Sequence 58, Application US/10124557
Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gessner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989

```

: APPLICATION NUMBER: US 07/390,901
: FILING DATE: 08-AUG-1989
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Geert, Luann
: REGISTRATION NUMBER: 31,822
: REFERENCE/DOCKET NUMBER: GI 5190
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 876-1170
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 58:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1049 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match          73.2%; Score 5510.5; DB 14; Length 1049;
Best Local Similarity 92.0%; Pred. No. 9.5e-273;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2

```

QY	1	MAMTLPYLLLLLSVPILOOVSSODLSSCAGRGEGVSRATCNCOVNCQHMECCPDF	60
Db	1	MAMTLPYLLLLLSVPILOOVSSQ-----	25
QY	61	KRVCTAELSCGRCFESFERGEDCDCAQCKKYDKCCPDYESFCAEVANPTSPSSKKAP	120
Db	26	-----ELSCGRCFESFERGEDCDCAQCKKYDKCCPDYSPCA-----	64
QY	121	PPGASQTIKSTYKRSPPKPNKKTKKVISEBETTEHVSVENOBSSSSSSSSSSSTIWW	180
Db	65	-----EEHVSVENOBSSSSSSSSSSSSSTIWW	89
QY	181	KIKSSKNSAANRELOKKLVXKONKKNPTKKKPTPKPVVDAGSGLDNGDKRYTTPDST	240
Db	90	KIKSSKNSAANRELOKKLVXKONKKNKKKPTPKPVVDAGSGLDNGDKRYTTPDST	149
QY	241	TOHNVKSTSPKITIYAKDINPRPSLPPNSDTSKETSLLTVNKEITYETKTTTNNKQTSIDG	300
Db	150	TOHNVKSTSPKITIYAKDINPRPSLPPNSDTSKETSLLTVNKEITYETKTTTNNKQTSIDG	209
QY	301	KEKTTSAKETOSIKTSAKODIAPTSKYLAKTPYAEITTKGPAITTPKEPPTTPKBPAS	360
Db	210	KEKTTSAKETOSIKTSAKODIAPTSKYLAKTPYAEITTKGPAITTPKEPPTTPKBPAS	269
QY	361	TTTKEBPPTTIKSAPTTPKBPAPTTTSSAPTTKBPAPTTTKEBPAPTTKBPAPTTKEP	420
Db	270	TTTKEBPPTTIKSAPTTPKBPAPTTTSSAPTTKBPAPTTTKEBPAPTTKBPAPTTKEP	329
QY	421	APTTSKAPPTPKBPAPPTPKBPAPTTPKBPAPTTPKBPAPTTTKEBPAPTTKBPAPTTPK	480
Db	330	APTTSKAPPTPKBPAPPTPKBPAPTTPKBPAPTTPKBPAPTTTKEBPAPTTKBPAPTTPK	389
QY	481	EPAPTPAKBPAPTTPKBPAPTTPKBPAPTTKBPAPTTKBPAPTTTSAPTTKEBAPVT	540
Db	390	EPAPTPAKBPAPTTPKBPAPTTPKBPAPTTKBPAPTTTSAPTTKEBAPVT	449
QY	541	TTTGSAPTTKBPAPTTTKEBPAPTTKBPAPTTKBPAPTTKBPAPTTTKEBPAPTTTKEP	600
Db	450	TTTGSAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP	509
QY	601	APTAPKBPAPPTPKETAPTTPPKULTPTTPEGLAFTTPEKBPAPTPPEELAPTTPEEPPTT	660
Db	510	APTAPKBPAPPTPKETAPTTPPKULTPTTPEGLAFTTPEKBPAPTPPEELAPTTPEEPPTT	569
QY	661	PEEBAPPTPKAABNTPKEBPAPTTPKBPAPTTPKBPAPTTPKETAPTTPKGTAPTTLKEP	720
Db	570	PEEBAPPTPKAABNTPKEBPAPTTPKBPAPTTPKBPAPTTPKETAPTTPKGTAPTTLKEP	629
QY	721	APTTPKBPAPKELAPTTTKEPTSTTSKBPAPTTKGPAPTTPKBPAPTTKBPAPTTPKG	780
Db	630	APTTPKBPAPKELAPTTTKEPTSTTSKBPAPTTKGPAPTTPKBPAPTTKBPAPTTPKG	689

QY 781 TAPTLKEBAPPTPKKPAKELAPTTKGPSTSTSDKPAPTPKETAPTTKKEBAPPTPK 840
 DB 690 TAPTLKEBAPPTPKKPAKELAPTTKGPSTSTSDKPAPTPKETAPTTKKEBAPPTPK 749
 QY 841 KCAPPTPETPTPTSTSVSTPTTKKEPTTIHKSPDESTPELSAPPTKALENSKPEGVPT 900
 DB 750 KCAPPTPETPTPTSTSVSTPTTKKEPTTIHKSPDESTPELSAPPTKALENSKPEGVPT 809
 QY 901 TTPPATKREMTTAAOKTTERDLRTTPPETTTAAPTOKETATTTTKESTKITATTTVOY 960
 DB 810 TTPPATKREMTTAAOKTTERDLRTTPPETTTAAPTOKETATTTTKESTKITATTTVOY 869
 QY 961 TTTTODTTPPKITLTKTTTLAPKVTTKKTIITTEIHKMPEETAPKORATNSKATTPK 1020
 DB 870 TTTTODTTPPKITLTKTTTLAPKVTTKKTIITTEIHKMPEETAPKORATNSKATTPK 929
 QY 1021 PCKPTAPKPKPTSTKPKTMPVRKPKTTPTPKMTSTWPELNPSTRIAEMLQTTTRN 1080
 DB 930 PCKPTAPKPKPTSTKPKTMPVRKPKTTPTPKMTSTWPELNPSTRIAEMLQTTTRN 989
 QY 1081 QTPNSKLVENPKSEBAGABEGTPTMLLRPHVFMPEVTPDMDYLRVFNQGIINPMIS 1140
 DB 990 QTPNSKLVENPKSEBAGABEGTPTMLLRPHVFMPEVTPDMDYLRVFNQGIINPMIS 1049

RESULT 14

US-10-124-557-84
 Sequence 84, Application US/10124557
 Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 Clark, Stephen C.

Jacobs, Kenneth
 Hewick, Rodney M.

Geener, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:

LENGTH: 1022 amino acids

TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 84:
 US-10-124-557-84

Query Match 72.1%; Score 5425.5; DB 14; Length 1022;
 Best Local Similarity 91.5%; Pred. No. 26-268;
 Matches 1020; Conservative 0; Mismatches 2; Indels 93; Gaps 2;

QY 26 DISSCAGRCGEGYSBDAATCNCDCYNQOHMECCPDFKRVCTABLSCKGRCSFBERGRED 85
 DB 1 DISSCAGRCGEGYSBDAATCNCDCYNQOHMECCPDFKRVCTABLSCKGRCSFBERGRED 58
 QY 86 CDAQCKKIDKCCPDYBESCAEYHNPSTSPSSKKAAPPAGSAGTISTTKRSFKPPNKKKT 145
 DB 59 CDAQCKKIDKCCPDYBESCAEYHNPSTSPSSKKAAPPAGSAGTISTTKRSFKPPNKKKT 80
 QY 146 KVISEETIEHSVENQESSSSSSSSSTIKIKSSKNSAANRELQKKLVKDNKK 205
 DB 81 -----AVKDNKK 87
 QY 206 NRTKKKPKPPEVNDAGSGLDNGDPKYTPDPTSTTOHNVKSTSPKITTAKBINRPSLP 265
 DB 88 NRTKKKPKPPEVNDAGSGLDNGDPKYTPDPTSTTOHNVKSTSPKITTAKBINRPSLP 147
 QY 266 PMSDTSKETSLSLVNKEITVETKETTINKQSTDKKETSNAKETSIEKTSAKDLAPTS 325
 DB 148 PMSDTSKETSLSLVNKEITVETKETTINKQSTDKKETSNAKETSIEKTSAKDLAPTS 207
 QY 326 KVLAKPTKAEITTTGPAITTPKEPTPTTPKESASTPPEPTTIKASAPTTKKEBAPT 385
 DB 208 KVLAKPTKAEITTTGPAITTPKEPTPTTPKESASTPPEPTTIKASAPTTKKEBAPT 267
 QY 386 TKSAPTTKKEBAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 445
 DB 268 TKSAPTTKKEBAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 327
 QY 446 TTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPT 505
 DB 328 TTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPT 387
 QY 506 PAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPT 565
 DB 388 PAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPT 447
 QY 566 KCAPPTPKKPAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 625
 DB 448 KCAPPTPKKPAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 507
 QY 626 PTPPEKLAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPT 685
 DB 508 PTPPEKLAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPT 567
 QY 686 KCAPPTPKKPAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 745
 DB 568 KCAPPTPKKPAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 627
 QY 746 SDKPAATTPKGAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 805
 DB 628 SDKPAATTPKGAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 687
 QY 806 TTKGPTSTTSOKPAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 865
 DB 688 TTKGPTSTTSOKPAATTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPKESAPTTPK 747
 QY 866 PTTIHKSPDESTPELSAPPTKALENSKPEGVPTTKPAATKPEMTTAAOKTTERDLR 925
 DB 748 PTTIHKSPDESTPELSAPPTKALENSKPEGVPTTKPAATKPEMTTAAOKTTERDLR 807
 QY 926 TTPETTTAAPTOKETATTTTKESTKITATTTVOYSTTTOOTTPFKITLTKTTLAPKV 985
 DB 808 TTPETTTAAPTOKETATTTTKESTKITATTTVOYSTTTOOTTPFKITLTKTTLAPKV 867

Qy	1141	DETNICGKPVDDGLTTLRNGTLVAFRGHYFMMLSPSPSPARITEVWGI	PSPIDTVFT	1200
Db	1007	-----	-----	1006
Qy	1201	RCNCEGKTFFPKDSQYWRFTNDIKDAGYPKPIFKFGGLTGQI	VAAI	STAKYNWPE
Db	1007	-----	-----	1006
Qy	1261	FFKRGGSIOQYIYKQEPVQKCPGRPALNYPVYGEMTOVRRRR	FERAIG	SGTHTIRIQY
Db	1007	-----	-----	1006
Qy	1321	SPARLAYQDKGYLHNEVKVSI	LRGLPNVVTSAISL	PNIRKPDGYDYAF
Db	1007	-----	-----	1014
Qy	1381	PSRTARAITTRSGQTL	SKWYNCP	1404
Db	1015	PSRTARAITTRSGQTL	SKWYNCP	1038

Search completed: December 8, 2003, 09:43:03
 Job time : 51 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:34:07 ; Search time 22 Seconds
(without alignments)
2700.202 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526

Sequence: 1 MAMKTLPIYLLLLSVFIQ.....ARAITRSGQTLSKWNCP 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

1: Issued Patents AA:
2: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6C_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	4	US-07-757-022B-2
2	7526	100.0	1404	4	US-07-757-022B-62
3	7291.5	96.9	1361	4	US-07-757-022B-40
4	7245.5	96.3	1363	4	US-07-757-022B-52
5	7232	96.1	1354	4	US-07-757-022B-48
6	7011	93.2	1320	4	US-07-757-022B-46
7	7011	93.2	1320	4	US-07-757-022B-42
8	7007.5	93.1	1311	4	US-07-757-022B-60
9	6971	92.6	1314	4	US-07-757-022B-50
10	6951.5	92.4	1313	4	US-07-757-022B-142
11	6727	89.4	1270	4	US-07-757-022B-44
12	6085	80.9	1140	4	US-07-757-022B-104
13	5510.5	73.2	1049	4	US-07-757-022B-58
14	5425.5	72.1	1022	4	US-07-757-022B-84
15	5335	70.9	1038	4	US-07-757-022B-74
16	5011	66.6	941	4	US-07-757-022B-154
17	2034.5	27.0	453	4	US-07-757-022B-54
18	1763.5	23.4	423	4	US-07-757-022B-66
19	1754	23.3	422	4	US-07-757-022B-68
20	1460	19.4	372	4	US-07-757-022B-96
21	1188	15.8	220	4	US-07-757-022B-94
22	1126	15.0	209	4	US-07-757-022B-132
23	1121	14.9	208	4	US-07-757-022B-116
24	1111.5	14.8	207	4	US-07-757-022B-136
25	1111.5	14.8	207	4	US-07-757-022B-92
26	1100	14.6	204	4	US-07-757-022B-70
27	1077	14.3	296	4	US-07-757-022B-70

28	1074	14.3	231	4	US-07-757-022B-30	Sequence 30, Appl
29	1041	13.8	192	4	US-07-757-022B-90	Sequence 90, Appl
30	945	12.6	172	4	US-07-757-022B-88	Sequence 88, Appl
31	881	11.7	188	4	US-07-757-022B-32	Sequence 32, Appl
32	871	11.6	156	4	US-07-757-022B-106	Sequence 106, App
33	871	11.6	156	4	US-07-757-022B-118	Sequence 118, App
34	827.5	11.0	157	4	US-07-757-022B-102	Sequence 102, App
35	827.5	11.0	157	4	US-07-757-022B-114	Sequence 114, App
36	806	10.7	217	4	US-07-757-022B-76	Sequence 76, Appl
37	800	10.6	141	4	US-07-757-022B-80	Sequence 80, Appl
38	750	10.0	131	4	US-07-757-022B-82	Sequence 82, Appl
39	743.5	9.9	237	4	US-07-757-022B-72	Sequence 72, Appl
40	740	9.8	130	4	US-07-757-022B-78	Sequence 78, Appl
41	740	9.8	130	4	US-07-757-022B-86	Sequence 86, Appl
42	706.5	9.4	132	4	US-07-757-022B-140	Sequence 140, App
43	702	9.3	8991	4	US-08-714-741-32	Sequence 32, Appl
44	618	8.2	138	4	US-07-757-022B-34	Sequence 34, Appl
45	613	8.1	111	4	US-07-757-022B-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-2

Query Match 100.0%; Score 7526; DB 4; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLILLISLVFIQOVSSODLSSCAGRCCEGYSRDATCNCDCVNCCHYMCCDF 60
DB 1 MAMKTLPIYLILLISLVFIQOVSSODLSSCAGRCCEGYSRDATCNCDCVNCCHYMCCDF 60
QY 61 KRVCIAEISCKRCCESEFERGECDDAOCKKYDKCCDYSEFCLEVNHPSPSSKAP 120
DB 61 KRVCIAEISCKRCCESEFERGECDDAOCKKYDKCCDYSEFCLEVNHPSPSSKAP 120
QY 61 KRVCIAEISCKRCCESEFERGECDDAOCKKYDKCCDYSEFCLEVNHPSPSSKAP 120
DB 61 KRVCIAEISCKRCCESEFERGECDDAOCKKYDKCCDYSEFCLEVNHPSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEIEIEHSVSNQSSSSSSSSSSSTIW 180
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEIEIEHSVSNQSSSSSSSSSSSTIW 180
QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEIEIEHSVSNQSSSSSSSSSSSTIW 180
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEIEIEHSVSNQSSSSSSSSSSSTIW 180
QY 181 KIKSSKNAANRELOKUKVNDKDKRTKKKPTPRPVVDAGSGLDNGDFKVTTPDST 240
DB 181 KIKSSKNAANRELOKUKVNDKDKRTKKKPTPRPVVDAGSGLDNGDFKVTTPDST 240
QY 181 KIKSSKNAANRELOKUKVNDKDKRTKKKPTPRPVVDAGSGLDNGDFKVTTPDST 240
DB 181 KIKSSKNAANRELOKUKVNDKDKRTKKKPTPRPVVDAGSGLDNGDFKVTTPDST 240
QY 241 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILYVNETVETKETTITNNKQTSIDG 300
DB 241 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILYVNETVETKETTITNNKQTSIDG 300
QY 241 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILYVNETVETKETTITNNKQTSIDG 300
DB 241 TOHNVSTSPKITTAKPINRPSLPNSDTSKETSILYVNETVETKETTITNNKQTSIDG 300
QY 301 KEKTSIAKETOSIEKTSADLAPTSIVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 360
DB 301 KEKTSIAKETOSIEKTSADLAPTSIVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 360
QY 301 KEKTSIAKETOSIEKTSADLAPTSIVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 360
DB 301 KEKTSIAKETOSIEKTSADLAPTSIVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 360
QY 361 TTPKEPTPTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEP 420
DB 361 TTPKEPTPTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEP 420
QY 361 TTPKEPTPTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEP 420
DB 361 TTPKEPTPTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEP 420
QY 421 APITTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTK 480
DB 421 APITTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTK 480
QY 421 APITTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTK 480
DB 421 APITTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTK 480
QY 481 EPAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 540
DB 481 EPAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 540
QY 481 EPAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 540
DB 481 EPAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 540
QY 541 TTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPAT 600
DB 541 TTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPAT 600
QY 541 TTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPAT 600
DB 541 TTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPAT 600
QY 601 APAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 660
DB 601 APAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 660
QY 601 APAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 660
DB 601 APAPAPKPPAPPTTKEBPATTTKSAATTTKEBPATTTKSAATTTKEBPATTTKEBPAT 660
QY 661 PEEBPATTTKAAAPNTPEKAPATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTKEP 720
DB 661 PEEBPATTTKAAAPNTPEKAPATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTKEP 720
QY 661 PEEBPATTTKAAAPNTPEKAPATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTKEP 720
DB 661 PEEBPATTTKAAAPNTPEKAPATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTKEP 720
QY 661 PEEBPATTTKAAAPNTPEKAPATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTKEP 720
DB 661 PEEBPATTTKAAAPNTPEKAPATTTKEBPATTTKSAATTTKEBPATTTKEBPATTTKEP 720
QY 721 APITTKKAPAPKELAPITTKKEPTSTSDKAPATTTKSAATTTKEBPATTTKEBPATTTKEP 780
DB 721 APITTKKAPAPKELAPITTKKEPTSTSDKAPATTTKSAATTTKEBPATTTKEBPATTTKEP 780
QY 721 APITTKKAPAPKELAPITTKKEPTSTSDKAPATTTKSAATTTKEBPATTTKEBPATTTKEP 780
DB 721 APITTKKAPAPKELAPITTKKEPTSTSDKAPATTTKSAATTTKEBPATTTKEBPATTTKEP 780
QY 781 TAPITTKKAPATTTKAPAPKELAPITTKGPTSTSDKAPATTTKSAATTTKEBPATTTKEP 840
DB 781 TAPITTKKAPATTTKAPAPKELAPITTKGPTSTSDKAPATTTKSAATTTKEBPATTTKEP 840
QY 781 TAPITTKKAPATTTKAPAPKELAPITTKGPTSTSDKAPATTTKSAATTTKEBPATTTKEP 840
DB 781 TAPITTKKAPATTTKAPAPKELAPITTKGPTSTSDKAPATTTKSAATTTKEBPATTTKEP 840
QY 841 KPAATTPETPTTSEVSTPTTKKEPTTIHNSPDSESTPELSABPTPKALENSPKPEGYPT 900
DB 841 KPAATTPETPTTSEVSTPTTKKEPTTIHNSPDSESTPELSABPTPKALENSPKPEGYPT 900
QY 841 KPAATTPETPTTSEVSTPTTKKEPTTIHNSPDSESTPELSABPTPKALENSPKPEGYPT 900
DB 841 KPAATTPETPTTSEVSTPTTKKEPTTIHNSPDSESTPELSABPTPKALENSPKPEGYPT 900
QY 901 TKTPAAKTPKEMTTAKDKTTERDRTATTPETTTAAFKMTKETATTTTEKTESKITAATTVOV 960
DB 901 TKTPAAKTPKEMTTAKDKTTERDRTATTPETTTAAFKMTKETATTTTEKTESKITAATTVOV 960
QY 901 TKTPAAKTPKEMTTAKDKTTERDRTATTPETTTAAFKMTKETATTTTEKTESKITAATTVOV 960
DB 901 TKTPAAKTPKEMTTAKDKTTERDRTATTPETTTAAFKMTKETATTTTEKTESKITAATTVOV 960
QY 961 TSTTTGDTTPPKITTLTKTTTLAPKVTTKITITTEINNKKEPTAKPRDRAATNSATTPK 1020
DB 961 TSTTTGDTTPPKITTLTKTTTLAPKVTTKITITTEINNKKEPTAKPRDRAATNSATTPK 1020

```

```

DB 961 TSTTTGDTTPPKITTLTKTTTLAPKVTTKITITTEINNKKEPTAKPRDRAATNSATTPK 1020
QY 1021 POKPTAPKPTSTYKPKETKMPVRKPKTTPPRKMTSTMPBLNPTSIAEAMLOTTREPN 1080
DB 1021 POKPTAPKPTSTYKPKETKMPVRKPKTTPPRKMTSTMPBLNPTSIAEAMLOTTREPN 1080
QY 1021 POKPTAPKPTSTYKPKETKMPVRKPKTTPPRKMTSTMPBLNPTSIAEAMLOTTREPN 1080
DB 1021 POKPTAPKPTSTYKPKETKMPVRKPKTTPPRKMTSTMPBLNPTSIAEAMLOTTREPN 1080
QY 1081 QTPNSKIVENVPKEDDGAAGETPHMLRPHVMEPTTPMDLPRVNGGIIINPMLS 1140
DB 1081 QTPNSKIVENVPKEDDGAAGETPHMLRPHVMEPTTPMDLPRVNGGIIINPMLS 1140
QY 1081 QTPNSKIVENVPKEDDGAAGETPHMLRPHVMEPTTPMDLPRVNGGIIINPMLS 1140
DB 1081 QTPNSKIVENVPKEDDGAAGETPHMLRPHVMEPTTPMDLPRVNGGIIINPMLS 1140
QY 1141 DETNINCKGPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITTEWGISPIDVFT 1200
DB 1141 DETNINCKGPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITTEWGISPIDVFT 1200
QY 1141 DETNINCKGPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITTEWGISPIDVFT 1200
DB 1141 DETNINCKGPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITTEWGISPIDVFT 1200
QY 1201 RCNCEGKTFPPKDSQYRFTNDIDAGPKPIFKFGGLTQIYVAALSTAKYKWPESVY 1260
DB 1201 RCNCEGKTFPPKDSQYRFTNDIDAGPKPIFKFGGLTQIYVAALSTAKYKWPESVY 1260
QY 1201 RCNCEGKTFPPKDSQYRFTNDIDAGPKPIFKFGGLTQIYVAALSTAKYKWPESVY 1260
DB 1201 RCNCEGKTFPPKDSQYRFTNDIDAGPKPIFKFGGLTQIYVAALSTAKYKWPESVY 1260
QY 1261 FFKRGGSIOQYIVKQEPVQKCPGRPALNYPVGEVMTQVRRRRERAIQSPQHTIRQY 1320
DB 1261 FFKRGGSIOQYIVKQEPVQKCPGRPALNYPVGEVMTQVRRRRERAIQSPQHTIRQY 1320
QY 1261 FFKRGGSIOQYIVKQEPVQKCPGRPALNYPVGEVMTQVRRRRERAIQSPQHTIRQY 1320
DB 1261 FFKRGGSIOQYIVKQEPVQKCPGRPALNYPVGEVMTQVRRRRERAIQSPQHTIRQY 1320
QY 1321 SPARLAYODKGVLENEVKVSIIMRGLPNVVTSAISLPIRKPDGYDYAFSKDQYINIDV 1380
DB 1321 SPARLAYODKGVLENEVKVSIIMRGLPNVVTSAISLPIRKPDGYDYAFSKDQYINIDV 1380
QY 1381 PSRTARAITTRSGQTLKSWYNCP 1404
DB 1381 PSRTARAITTRSGQTLKSWYNCP 1404
QY 1381 PSRTARAITTRSGQTLKSWYNCP 1404
DB 1381 PSRTARAITTRSGQTLKSWYNCP 1404

```

RESULT 2
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Geener, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserti, Luann

REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 62:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULAR TYPE: protein
 US-07-757-022B-62

Query Match 100.0%; Score 7526; DB 4; Length 1404;
 Basic Local Similarity 100.0%; Pred. No. 0;
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAMKTLPIYLLLSVFIQVSSODLSGAGRGEGSRDATCNCNDYNOHYMECCPDF 60
DB 1 MAMKTLPIYLLLSVFIQVSSODLSGAGRGEGSRDATCNCNDYNOHYMECCPDF 60
QY 61 KAVCTBELSCGKCFESFERGECDDAQCCKTDCCPDYESTCAVHNPTSPSSKAP 120
DB 61 KAVCTBELSCGKCFESFERGECDDAQCCKTDCCPDYESTCAVHNPTSPSSKAP 120
QY 121 PPSGASQTIKSTTKSPKPNKKTKKVISESEITEHVSSENOSSSSSSSSSTIM 180
DB 121 PPSGASQTIKSTTKSPKPNKKTKKVISESEITEHVSSENOSSSSSSSSSTIM 180
QY 181 KIKSSKNSANRELQKKLVKONKKNRTKKPTPKPVVDGASGLDNGDFKVTTPDST 240
DB 181 KIKSSKNSANRELQKKLVKONKKNRTKKPTPKPVVDGASGLDNGDFKVTTPDST 240
QY 241 TQHNKISTGPKITTAQINPRLPNSDTSKETSITVANKETVEKETTINKQTSIDG 300
DB 241 TQHNKISTGPKITTAQINPRLPNSDTSKETSITVANKETVEKETTINKQTSIDG 300
QY 301 KKKTTSAKETSOSIEKTSANDLAPTSKVLAKPPKATTTTGGPALTTPKEPTTPKEPAS 360
DB 301 KKKTTSAKETSOSIEKTSANDLAPTSKVLAKPPKATTTTGGPALTTPKEPTTPKEPAS 360
QY 361 TTPKEPTPTTIKSAPTTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPAATTTKEP 420
DB 361 TTPKEPTPTTIKSAPTTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPAATTTKEP 420
QY 421 APTTTSAPTTPEKPAATTPPKPAATTPKEPAATTTPEPTTPKEPAATTPKEPAATTPK 480
DB 421 APTTTSAPTTPEKPAATTPPKPAATTPKEPAATTTPEPTTPKEPAATTPKEPAATTPK 480
QY 481 BPAPAPKPAATTPKEPAATTPKEPAATTTKESPTTPKEPAATTTKSAATTPKEPAAT 540
DB 481 BPAPAPKPAATTPKEPAATTPKEPAATTTKESPTTPKEPAATTTKSAATTPKEPAAT 540
QY 541 TTKSAPTTPEKSPPTTKSAATTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPAATTTKEP 600
DB 541 TTKSAPTTPEKSPPTTKSAATTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPAATTTKEP 600
QY 541 TTKSAPTTPEKSPPTTKSAATTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPAATTTKEP 600
DB 541 TTKSAPTTPEKSPPTTKSAATTPKEPAATTTKSAATTPKEPAATTTKSAATTPKEPAATTTKEP 600
QY 601 AAPAPKEPAATTPKEPAATTPPKKLTPTTBEKLAPTTPEKPAATTPBEELAPTTPEEPTPT 660
DB 601 AAPAPKEPAATTPKEPAATTPPKKLTPTTBEKLAPTTPEKPAATTPBEELAPTTPEEPTPT 660
QY 661 PEEPAATTTKAAAPNTPEKPAATTPKEPAATTPKEPAATTTKETAATTTKGAATTTLKRP 720
DB 661 PEEPAATTTKAAAPNTPEKPAATTPKEPAATTPKEPAATTTKETAATTTKGAATTTLKRP 720
QY 721 AAPTPPKKPAKELAPTTKEPTSTSDKAPPTPKSTAATTPKEPAATTPKEPAATTPKG 780
DB 721 AAPTPPKKPAKELAPTTKEPTSTSDKAPPTPKSTAATTPKEPAATTPKEPAATTPKG 780
QY 781 TAPTTLKPAATTPPKKPAKELAPTTTGGTSTTSKAPATTPKEPAATTPKEPAATTPK 840
DB 781 TAPTTLKPAATTPPKKPAKELAPTTTGGTSTTSKAPATTPKEPAATTPKEPAATTPK 840

```

```

QY 841 KCAPTTPETPPTTSEVSTPTTKETTHKSPDSTPELSAAPTPEKALENSPEKGVPT 900
DB 841 KCAPTTPETPPTTSEVSTPTTKETTHKSPDSTPELSAAPTPEKALENSPEKGVPT 900
QY 901 TKTPAATKEMTTAKDKTERDLRTTPETTAAAPMTKETAATTEKTESKITATTTQV 960
DB 901 TKTPAATKEMTTAKDKTERDLRTTPETTAAAPMTKETAATTEKTESKITATTTQV 960
QY 961 TSTTODTTPPKITTLTKTTTLAPKVTYTTKKITTTTIEINAKPEETAKKORATNSKATTPK 1020
DB 961 TSTTODTTPPKITTLTKTTTLAPKVTYTTKKITTTTIEINAKPEETAKKORATNSKATTPK 1020
QY 1021 POKPTKAPKPKPTSTKPKTPMVRKPKTPTPRKMTSTMPBELNPTSRILEAMLOTTTRPN 1080
DB 1021 POKPTKAPKPKPTSTKPKTPMVRKPKTPTPRKMTSTMPBELNPTSRILEAMLOTTTRPN 1080
QY 1081 QTPNSKLVFNPKSDAGAGEBTPHMLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1140
DB 1081 QTPNSKLVFNPKSDAGAGEBTPHMLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1140
QY 1141 DETNLCNGKPVUGLTLLRNGTLVAFRGHYFMMLSPSPSPARRITTEWGISPIDTVTT 1200
DB 1141 DETNLCNGKPVUGLTLLRNGTLVAFRGHYFMMLSPSPSPARRITTEWGISPIDTVTT 1200
QY 1201 RCNCEGKTFEFDOSQYMRPTNDIKDAGYKPIFKGFGGLTGOIYVALSTAKYKNPESVY 1260
DB 1201 RCNCEGKTFEFDOSQYMRPTNDIKDAGYKPIFKGFGGLTGOIYVALSTAKYKNPESVY 1260
QY 1261 PFKGGSIIQOYLYKQEPVQKCPGRPALNYPYGEMTQVRRRRFERALGPSQTHIRIQY 1320
DB 1261 PFKGGSIIQOYLYKQEPVQKCPGRPALNYPYGEMTQVRRRRFERALGPSQTHIRIQY 1320
QY 1321 SPARLAYODKGVLANHEVKVSIIMRGLPNVVTSAISLPIRKDGIDYAFSSQOYINIDV 1380
DB 1321 SPARLAYODKGVLANHEVKVSIIMRGLPNVVTSAISLPIRKDGIDYAFSSQOYINIDV 1380
QY 1381 PSRTARAITTSRGOTLSKVNCP 1404
DB 1381 PSRTARAITTSRGOTLSKVNCP 1404

```

RESULT 3
 US-07-757-022B-40
 ; Sequence 40, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geier, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-40

Query Match 96.9%; Score 7291.5; DB 4; Length 1361;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAMKTLPIYLILLISVFYIQOVSQDLSGCGRGEGYSDATGNCDCYNCHVMECCPDF 60
DB 1 MAMKTLPIYLILLISVFYIQOVSQDLSGCGRGEGYSDATGNCDCYNCHVMECCPDF 60
QY 61 KRVTAEIASCGRCESEFERGECDCDAQCKKYDCCPDYSEFCAEVNPSTSPSSKAP 120
DB 61 KRVTAEIASCGRCESEFERGECDCDAQCKKYDCCPDYSEFCAEVNPSTSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSKPRPKKTKTKVIESBEITEHSVSENQSSSSSSSSSTIw 180
DB 121 PPSGASQTIKSTTKRSKPRPKKTKTKVIESBEITEHSVSENQSSSSSSSSSTIw 180
QY 121 PPSGASQTIKSTTKRSKPRPKKTKTKVIESBEITEHSVSENQSSSSSSSSSTIw 156
DB 121 PPSGASQTIKSTTKRSKPRPKKTKTKVIESBEITEHSVSENQSSSSSSSSSTIw 156
QY 181 KIKSKNSAARELOKLVKNDKKNRTKKKTPKPVVDEAGSLDNGDFVTTPTST 240
DB 181 KIKSKNSAARELOKLVKNDKKNRTKKKTPKPVVDEAGSLDNGDFVTTPTST 240
QY 157 -----VDNKKNRKTKKTPKPVVDEAGSLDNGDFVTTPTST 197
DB 157 -----VDNKKNRKTKKTPKPVVDEAGSLDNGDFVTTPTST 197
QY 241 TQHNKVSSTPKITTAAPINPRPSLPNSDTSKETSLLVANKETVETKETTNNKOTSIDG 300
DB 241 TQHNKVSSTPKITTAAPINPRPSLPNSDTSKETSLLVANKETVETKETTNNKOTSIDG 300
QY 198 TQHNKVSSTPKITTAAPINPRPSLPNSDTSKETSLLVANKETVETKETTNNKOTSIDG 257
DB 198 TQHNKVSSTPKITTAAPINPRPSLPNSDTSKETSLLVANKETVETKETTNNKOTSIDG 257
QY 301 KEKTTSAETOSIEKTSKDLAPTSKVLAKTPPKAETTTKGPALTTPKEPTTPPKEPAS 360
DB 301 KEKTTSAETOSIEKTSKDLAPTSKVLAKTPPKAETTTKGPALTTPKEPTTPPKEPAS 360
QY 258 KEKTTSAETOSIEKTSKDLAPTSKVLAKTPPKAETTTKGPALTTPKEPTTPPKEPAS 317
DB 258 KEKTTSAETOSIEKTSKDLAPTSKVLAKTPPKAETTTKGPALTTPKEPTTPPKEPAS 317
QY 361 TTPKEPTTTIKSAPTTKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEP 420
DB 361 TTPKEPTTTIKSAPTTKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEP 420
QY 318 TTPKEPTTTIKSAPTTKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEP 377
DB 318 TTPKEPTTTIKSAPTTKEBPATTTKSAPTTPKEBPATTTKEBPATTTKEBPATTTKEP 377
QY 421 APTTTSAPTTKEBPATTTKPKAPTTTPKEBPATTTTPKEBPATTTKEBPATTTK 480
DB 421 APTTTSAPTTKEBPATTTKPKAPTTTPKEBPATTTTPKEBPATTTTPKEBPATTTK 480
QY 378 APTTTSAPTTKEBPATTTKPKAPTTTPKEBPATTTTPKEBPATTTTPKEBPATTTK 437
DB 378 APTTTSAPTTKEBPATTTKPKAPTTTPKEBPATTTTPKEBPATTTTPKEBPATTTK 437
QY 481 EPAPAPKAPATTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 540
DB 481 EPAPAPKAPATTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 540
QY 438 EPAPAPKAPATTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 497
DB 438 EPAPAPKAPATTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 497
QY 541 TTKSAPTTPKESPPTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 600
DB 541 TTKSAPTTPKESPPTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 600
QY 498 TTKSAPTTPKESPPTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 557
DB 498 TTKSAPTTPKESPPTTKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 557
QY 601 APTAPKEBPATTTPKETAFTTPKCLTPTPEKLAFTTPKEBPATTTPEBEAPTTPEBEPTTT 660
DB 601 APTAPKEBPATTTPKETAFTTPKCLTPTPEKLAFTTPKEBPATTTPEBEAPTTPEBEPTTT 660
QY 558 APTAPKEBPATTTPKETAFTTPKCLTPTPEKLAFTTPKEBPATTTPEBEAPTTPEBEPTTT 617
DB 558 APTAPKEBPATTTPKETAFTTPKCLTPTPEKLAFTTPKEBPATTTPEBEAPTTPEBEPTTT 617
QY 661 PEEBAPTTPKAAAPTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 720
DB 661 PEEBAPTTPKAAAPTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 720
QY 618 PEEBAPTTPKAAAPTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 677
DB 618 PEEBAPTTPKAAAPTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPAT 677

QY 721 APTTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPKPAFTTPK 780
DB 721 APTTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPKPAFTTPK 780
QY 678 APTTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPKPAFTTPK 737
DB 678 APTTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPKPAFTTPK 737
QY 781 TAPTTLKEBPATTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPK 840
DB 781 TAPTTLKEBPATTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPK 840
QY 738 TAPTTLKEBPATTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPK 797
DB 738 TAPTTLKEBPATTPKPAKELAPTTTKEPTSTSDKAPATTPKGAFTTPKKEBPATTPK 797
QY 841 KPAFTTPEPTTSEVSTPTTKEPTTIKSPDESIPBELSAEPTPALENSPEKPGPT 900
DB 841 KPAFTTPEPTTSEVSTPTTKEPTTIKSPDESIPBELSAEPTPALENSPEKPGPT 900
QY 798 KPAFTTPEPTTSEVSTPTTKEPTTIKSPDESIPBELSAEPTPALENSPEKPGPT 857
DB 798 KPAFTTPEPTTSEVSTPTTKEPTTIKSPDESIPBELSAEPTPALENSPEKPGPT 857
QY 901 TKTPAATPEMTTAKKOTTERDRTTPEPTTAAPKMTKETATTTTEKTESKITATTTQV 960
DB 901 TKTPAATPEMTTAKKOTTERDRTTPEPTTAAPKMTKETATTTTEKTESKITATTTQV 960
QY 858 TKTPAATPEMTTAKKOTTERDRTTPEPTTAAPKMTKETATTTTEKTESKITATTTQV 917
DB 858 TKTPAATPEMTTAKKOTTERDRTTPEPTTAAPKMTKETATTTTEKTESKITATTTQV 917
QY 961 TSTTTQDTPPKITTLTKITTLAPRVTTTKITITTEIMNKBEETAKPKDRATNSKATTPK 1020
DB 961 TSTTTQDTPPKITTLTKITTLAPRVTTTKITITTEIMNKBEETAKPKDRATNSKATTPK 1020
QY 918 TSTTTQDTPPKITTLTKITTLAPRVTTTKITITTEIMNKBEETAKPKDRATNSKATTPK 977
DB 918 TSTTTQDTPPKITTLTKITTLAPRVTTTKITITTEIMNKBEETAKPKDRATNSKATTPK 977
QY 1021 POKETKAPKSTSTTKKPTMPRVKPKTTPPRKOTSTMPBLNPTSRIAEAMLQTTTRPN 1080
DB 1021 POKETKAPKSTSTTKKPTMPRVKPKTTPPRKOTSTMPBLNPTSRIAEAMLQTTTRPN 1080
QY 978 POKETKAPKSTSTTKKPTMPRVKPKTTPPRKOTSTMPBLNPTSRIAEAMLQTTTRPN 1037
DB 978 POKETKAPKSTSTTKKPTMPRVKPKTTPPRKOTSTMPBLNPTSRIAEAMLQTTTRPN 1037
QY 1081 QTPNSKIVEVNPKESEDAGAGETPHMLRPHVFMPEKTPMDYLPVPNOGIIINPMLS 1140
DB 1081 QTPNSKIVEVNPKESEDAGAGETPHMLRPHVFMPEKTPMDYLPVPNOGIIINPMLS 1140
QY 1038 QTPNSKIVEVNPKESEDAGAGETPHMLRPHVFMPEKTPMDYLPVPNOGIIINPMLS 1097
DB 1038 QTPNSKIVEVNPKESEDAGAGETPHMLRPHVFMPEKTPMDYLPVPNOGIIINPMLS 1097
QY 1141 DETNINCKGPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITTEWIGISPDIYFT 1200
DB 1141 DETNINCKGPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITTEWIGISPDIYFT 1200
QY 1098 DETNINCKGPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITTEWIGISPDIYFT 1157
DB 1098 DETNINCKGPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITTEWIGISPDIYFT 1157
QY 1201 RNCCEGTFEPKDSQVRFMTNDIDAGYPKIIFGFGGLTQIYAAIATAKYKMPBSVY 1260
DB 1201 RNCCEGTFEPKDSQVRFMTNDIDAGYPKIIFGFGGLTQIYAAIATAKYKMPBSVY 1260
QY 1158 RNCCEGTFEPKDSQVRFMTNDIDAGYPKIIFGFGGLTQIYAAIATAKYKMPBSVY 1217
DB 1158 RNCCEGTFEPKDSQVRFMTNDIDAGYPKIIFGFGGLTQIYAAIATAKYKMPBSVY 1217
QY 1261 PFKRGSIQOYIYQOEPQOKCPGRPALNYPVYGMTOVRRRRRERAIQPSQTHIRIQY 1320
DB 1261 PFKRGSIQOYIYQOEPQOKCPGRPALNYPVYGMTOVRRRRRERAIQPSQTHIRIQY 1320
QY 1218 PFKRGSIQOYIYQOEPQOKCPGRPALNYPVYGMTOVRRRRRERAIQPSQTHIRIQY 1277
DB 1218 PFKRGSIQOYIYQOEPQOKCPGRPALNYPVYGMTOVRRRRRERAIQPSQTHIRIQY 1277
QY 1321 SPARLAYQDKVILNEKVSILMRGLPNVVTSAISLPIKRPDGYDYAFSKQYINIDV 1380
DB 1321 SPARLAYQDKVILNEKVSILMRGLPNVVTSAISLPIKRPDGYDYAFSKQYINIDV 1380
QY 1278 SPARLAYQDKVILNEKVSILMRGLPNVVTSAISLPIKRPDGYDYAFSKQYINIDV 1337
DB 1278 SPARLAYQDKVILNEKVSILMRGLPNVVTSAISLPIKRPDGYDYAFSKQYINIDV 1337
QY 1381 PSRTARAITTRSGQTLKRWYNCP 1404
DB 1381 PSRTARAITTRSGQTLKRWYNCP 1404
QY 1338 PSRTARAITTRSGQTLKRWYNCP 1361
DB 1338 PSRTARAITTRSGQTLKRWYNCP 1361

RESULT 4
US-07-757-022B-52
Sequence 52, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cheer, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-52

Query Match 96.3%; Score 7245.5; DB 4; Length 1363;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAMKTLPIYLLLSVFIQVSSODLSGCAGRCGEGYSDATCNCDCYNCOHYMECCPDE 60
DB 1 MAMKTLPIYLLLSVFIQVSSODLSGCAGRCGEGYSDATCNCDCYNCOHYMECCPDE 25
QY 61 KAVCTAELSCGKGFSPERGRCDDAOCKKTDKCCPDYESTCAVYHNTSPSSKAP 120
DB 26 -----ELSCGKGFSPERGRCDDAOCKKTDKCCPDYESTCAVYHNTSPSSKAP 79
QY 121 PRGSAQTIKSTTKRSPKPKKKTKVISEEITEBSHVSSENOSSSSSSSSSSSTTM 180
DB 80 PRGSAQTIKSTTKRSPKPKKKTKVISEEITEBSHVSSENOSSSSSSSSSSSTTM 139
QY 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPVVDEAGSGLDNGDFKVTPTDST 240
DB 140 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPVVDEAGSGLDNGDFKVTPTDST 199
QY 241 TOHNKSTSPKITTAKPINDRPSLPNSDTSKETSLSLVNKEITVEKETTNNKQSTDG 300
DB 200 TOHNKSTSPKITTAKPINDRPSLPNSDTSKETSLSLVNKEITVEKETTNNKQSTDG 259
QY 301 KKKTSKAKETOSIEKTSANDLAFSTSVLAKPTKAEITTKGPAITTKPKBPPTPKBPAS 360
DB 260 KKKTSKAKETOSIEKTSANDLAFSTSVLAKPTKAEITTKGPAITTKPKBPPTPKBPAS 319
QY 361 TTPKEBPTTIKSAPTTPKEBAPTITTKSAPTTPKEBAPTITTKSAPTTPKEBAPTITTK 420
DB 320 TTPKEBPTTIKSAPTTPKEBAPTITTKSAPTTPKEBAPTITTKSAPTTPKEBAPTITTK 379
QY 421 APTTTSAPTTPKEBAPTTPKAPPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 480
DB 380 APTTTSAPTTPKEBAPTTPKAPPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 439
QY 481 BPAATPKKAPPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 540
DB 440 BPAATPKKAPPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 499
QY 541 TTKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 600

DB 500 TTKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKAPPTTPKEBAPTTPKEBAPTTPK 559
QY 601 APTAPKEBAPTTPKETAAPTTPKLTPTTPEKLAFTTPKECAPTTPBELAFTTPBEPTPT 660
DB 560 APTAPKEBAPTTPKETAAPTTPKLTPTTPEKLAFTTPKECAPTTPBELAFTTPBEPTPT 619
QY 661 PEEBAPTTPKAAAPNTPKBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 720
DB 620 PEEBAPTTPKAAAPNTPKBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 679
QY 721 APTTPKEBAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKEBAPTTPKEBAPTTPK 780
DB 680 APTTPKEBAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKEBAPTTPKEBAPTTPK 739
QY 781 TAPTTKEBAPTTPKAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKEBAPTTPKEBAPT 840
DB 740 TAPTTKEBAPTTPKAPKELAPTTTKEPTSTSDKAPPTPKGTAPTTPKEBAPTTPKEBAPT 799
QY 841 KAPPTTETPTTSEVSTPTTKEBPTTIHKSPEDESTPELSAAPTPEKALENSPKBQVPT 900
DB 800 KAPPTTETPTTSEVSTPTTKEBPTTIHKSPEDESTPELSAAPTPEKALENSPKBQVPT 859
QY 901 TKTPAATKEEMTTTAKDKTTERDLRTTPTTAAAPMTKETATTTTEKTESKITATTTQV 960
DB 860 TKTPAATKEEMTTTAKDKTTERDLRTTPTTAAAPMTKETATTTTEKTESKITATTTQV 919
QY 961 TSTTTODTTPFKITTLKTTTLAPKVTYTTKKTITTEINAKPBETAPKORATNSKATTPK 1020
DB 920 TSTTTODTTPFKITTLKTTTLAPKVTYTTKKTITTEINAKPBETAPKORATNSKATTPK 979
QY 1021 POKPTKAPKKPTSTKKPKTPMVRKPKTPTPRKMTSTMPBELNPTSRILAEMLQTTTRN 1080
DB 980 POKPTKAPKKPTSTKKPKTPMVRKPKTPTPRKMTSTMPBELNPTSRILAEMLQTTTRN 1039
QY 1081 QTPNSKLVFNPKSEBAGABETPHMLLRPHVFMPEVTPDMDYLRPVNOGIIINPMLS 1140
DB 1040 QTPNSKLVFNPKSEBAGABETPHMLLRPHVFMPEVTPDMDYLRPVNOGIIINPMLS 1099
QY 1141 DETNTCNGKPVNGLTTLRNGTILVAPRGHFMMLSPSPSPSPARITTEWGISPTVTPT 1200
DB 1100 DETNTCNGKPVNGLTTLRNGTILVAPRGHFMMLSPSPSPSPARITTEWGISPTVTPT 1159
QY 1201 RCNCEGTFFFDOSQWRFTNDIKDAGYKPIFKGFGSLTGOIIVALASTAKKNMPEASY 1260
DB 1160 RCNCEGTFFFDOSQWRFTNDIKDAGYKPIFKGFGSLTGOIIVALASTAKKNMPEASY 1219
QY 1261 PFKGGSIOQYIYKQEPVQKCPGRPALNYPYVGEMTOVRRRRFERALGPSQHTIRIOY 1320
DB 1220 PFKGGSIOQYIYKQEPVQKCPGRPALNYPYVGEMTOVRRRRFERALGPSQHTIRIOY 1279
QY 1321 SPARLAYQDKVLAHNEVKVSIIMRGLPNVVTSAISLPINRKPDGYDYAFSQQYVNDIV 1380
DB 1280 SPARLAYQDKVLAHNEVKVSIIMRGLPNVVTSAISLPINRKPDGYDYAFSQQYVNDIV 1339
QY 1381 PSTARAITTRSGOTLSKWNVNC 1404
DB 1340 PSTARAITTRSGOTLSKWNVNC 1363

RESULT 5
US-07-757-022B-48
Sequence 48. Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Genent, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive

CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-5851
TELEPHONE: (617)876-1170
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-48

Query Match 96.1%; Score 7232; DB 4; Length 1354;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 1354; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAMKTLPIYLLLSVFIIQVSSODLSSCAGRCGEGRSDATCNCDCNQCCHWMECCDF 60
DB 1 MAMKTLPIYLLLSVFIIQVSSODLSSCAGRCGEGRSDATCNCDCNQCCHWMECCDF 60
QY 61 KRVCTAELSCGRCPESFERGREGCCDAQCKKYDKCCPDYESFCAEVNPTSPSSKAP 120
DB 61 KRVCTAELSCGRCPESFERGREGCCDAQCKKYDKCCPDYESFCAEVNPTSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPKKKTKKVI ESEBITHEHSVSENDESSSSSSSSSTIW 180
DB 106 ----- EHSVSENDESSSSSSSSSTIW 130
QY 181 KIKSSKSAARELOKLVKNDKKNRTKKKPTPEPVVDAGSLDNGDKVTTPDST 240
DB 131 KIKSSKSAARELOKLVKNDKKNRTKKKPTPEPVVDAGSLDNGDKVTTPDST 190
QY 241 TQHNKVSPTITKAPINRPSLPNSDTSKETSLVANKETVETKETTNNKQSTDG 300
DB 191 TQHNKVSPTITKAPINRPSLPNSDTSKETSLVANKETVETKETTNNKQSTDG 250
QY 301 KEKTTSAKETOSIEKTSADIAPTSIVLAKPTPKAETTTKGPALTTPEPTTPKEPAS 360
DB 251 KEKTTSAKETOSIEKTSADIAPTSIVLAKPTPKAETTTKGPALTTPEPTTPKEPAS 310
QY 361 TTPKEPTPTTITKSAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 420
DB 311 TTPKEPTPTTITKSAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 370
QY 421 APTTTSAPTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEK 480

DB 371 APTTTSAPTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEK 430
QY 481 EPAPLAPKKAPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTKSAPTTKEBPAT 540
DB 431 EPAPLAPKKAPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTKSAPTTKEBPAT 490
QY 541 TTKSAPTTPEKBPATTTKEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTKEP 600
DB 491 TTKSAPTTPEKBPATTTKEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTKEP 550
QY 601 APTAPKEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTT 660
DB 551 APTAPKEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTT 610
QY 661 PEEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEK 720
DB 611 PEEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEK 670
QY 721 APTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTT 780
DB 671 APTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTT 730
QY 781 TAPTTKEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTT 840
DB 731 TAPTTKEBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTTPEKBPATTT 790
QY 841 KPAATTPETPTTSEVSTPTTKEPTTIKSDPESTPELSABPTPALENSPEKPGVPT 900
DB 791 KPAATTPETPTTSEVSTPTTKEPTTIKSDPESTPELSABPTPALENSPEKPGVPT 850
QY 901 TKTPAATPEMTTAKOTTERDRTTETTAAPKMTKETATTEKTESKIRATTTQV 960
DB 851 TKTPAATPEMTTAKOTTERDRTTETTAAPKMTKETATTEKTESKIRATTTQV 910
QY 961 TSTTODTTPKITTLLKTTTLAPRVTTTKIITTEIMNKEETAKDRAATNSKATTPK 1020
DB 911 TSTTODTTPKITTLLKTTTLAPRVTTTKIITTEIMNKEETAKDRAATNSKATTPK 970
QY 1021 POKPTAKPKKTSIKKKTMRVRVKKPTTPPRKMTSTMPRLNTSRIAEAMLQTTTRPN 1080
DB 971 POKPTAKPKKTSIKKKTMRVRVKKPTTPPRKMTSTMPRLNTSRIAEAMLQTTTRPN 1030
QY 1081 QTPNSKIVEVNPKEDEGAGETPHMLRPHVMPETTPMDYLPRVNOGIIINPMLS 1140
DB 1031 QTPNSKIVEVNPKEDEGAGETPHMLRPHVMPETTPMDYLPRVNOGIIINPMLS 1090
QY 1141 DETNINCNGKPYDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPIDTVFT 1200
DB 1091 DETNINCNGKPYDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPIDTVFT 1150
QY 1201 RCNCEGKTFPPKQSVYRFTNDIDAGYKPIIFGFGGLGQIYIAALSTAYKXWPPSVY 1260
DB 1151 RCNCEGKTFPPKQSVYRFTNDIDAGYKPIIFGFGGLGQIYIAALSTAYKXWPPSVY 1210
QY 1261 FEKRGGSIQOYIYQOEVOVKOPGRPALNYPVGEEMQVRRRRERAIIGSQHTTIRIQY 1320
DB 1211 FEKRGGSIQOYIYQOEVOVKOPGRPALNYPVGEEMQVRRRRERAIIGSQHTTIRIQY 1270
QY 1321 SPARLAQDQGVLANEVKSLMRGLPNVYTSALSLNIRKPDGYDYVAFSKQDYNNIDV 1380
DB 1271 SPARLAQDQGVLANEVKSLMRGLPNVYTSALSLNIRKPDGYDYVAFSKQDYNNIDV 1330
QY 1381 PSRTARAITTRSGQTLISKWYNCP 1404
DB 1331 PSRTARAITTRSGQTLISKWYNCP 1354

RESULT 6
US-07-757-022B-46
; Sequence 46, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:

APPLICANT: Geener, Thomas G.
 APPLICANT: Clark, Stephen C.
 APPLICANT: Turner, Katherine
 APPLICANT: Hewick, Rodney M.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SORTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Geier, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 46:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1320 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-46

Query Match 93.2%; Score 7011; DB 4; Length 1320;
 Best Local Similarity 94.0%; Pctd. No. 0;
 Matches 1340; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

1 MAMKTLPIYLLLLSVFVIQOVSSQDLSGAGRGEGYRDATCNCNQCQHYMECCPDF 60
 1 MAMKTLPIYLLLLSVFVIQOVSSQ----- 25
 61 KAVCTAELSCGKCFESFERGRCDCDAOCKYDKCCPYESFCAEVNPTSPSSKAP 120
 26 -----ELSCGKCFESFERGRCDCDAOCKYDKCCPYESFCAEVNPTSPSSKAP 79
 121 PPSGASQTIKSTTKRSPKPKPKKKTKKVI ESEETIEHSEVSEHQESSSSSSSSSSSTIW 180
 80 PPSGASQTIKSTTKRSPKPKPKKKTKKVI ESEETIEHSEVSEHQESSSSSSSSSSSSSTIW 115
 181 KIKSKNSANRELQKKLVKONKKKRTKKPTPKPPVVD EAGSGLDNDPFKVTTPDTST 240
 116 -----VKONKKRTKKKPTPKPPVVD EAGSGLDNDPFKVTTPDTST 156
 241 TQHNKVSSTPKITTAAPINRPSLPNSDTSKETSITVNKETTVEETKTTTNKQSTDG 300
 157 TQHNKVSSTPKITTAAPINRPSLPNSDTSKETSITVNKETTVEETKTTTNKQSTDG 216

301 KERTTSKETSOSIEKTSADKLAPTSKVLAKPKAETTTKGPALTTPKEPTTPKEPS 360
 217 KERTTSKETSOSIEKTSADKLAPTSKVLAKPKAETTTKGPALTTPKEPTTPKEPS 276
 361 TTPKEPTTPKISAPTTKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420
 277 TTPKEPTTPKISAPTTKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 336
 421 APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 480
 337 APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 396
 481 BPAPTKAPKAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 540
 397 BPAPTKAPKAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 456
 541 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600
 457 TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 516
 601 APTAKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 660
 517 APTAKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 576
 661 PBPAPTTPKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 720
 577 PBPAPTTPKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 636
 721 APTPKKAPKELAPTTKBPAPTTSDKAPPTPKGIAATTPKBPAPTTKBPAPTTK 780
 637 APTPKKAPKELAPTTKBPAPTTSDKAPPTPKGIAATTPKBPAPTTKBPAPTTK 696
 781 TAPTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTSDKAPPTPKGIAATTPKBP 840
 697 TAPTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTSDKAPPTPKGIAATTPKBP 756
 841 KPAPTTPEPTTSEVSTPTTKBPAPTTKBPAPTTSDKAPPTPKGIAATTPKBP 900
 757 KPAPTTPEPTTSEVSTPTTKBPAPTTKBPAPTTSDKAPPTPKGIAATTPKBP 816
 901 TKTPAATKEMTTAKDKTERDLRTTETTTAABKMTETATTEKTESKITATTTTQV 960
 817 TKTPAATKEMTTAKDKTERDLRTTETTTAABKMTETATTEKTESKITATTTTQV 876
 961 TSTTQDTTPPKITTLKTTTLAPKVTYTTKITTETIENKPEETAKKORATNSKATPK 1020
 877 TSTTQDTTPPKITTLKTTTLAPKVTYTTKITTETIENKPEETAKKORATNSKATPK 936
 1021 POKPTKAPKKPTSTKKPKTPMPVRKPKTTPPRKMTSTMPBLNPTSTIAEMLQTTTRN 1080
 937 POKPTKAPKKPTSTKKPKTPMPVRKPKTTPPRKMTSTMPBLNPTSTIAEMLQTTTRN 996
 1081 QTPNSKLVEVNPSESDAGABGETPHMLLRPHVFEVTPDMVDYLPRVNOGIIINPMS 1140
 997 QTPNSKLVEVNPSESDAGABGETPHMLLRPHVFEVTPDMVDYLPRVNOGIIINPMS 1056
 1141 DETNINCGKPVDTGLTTLNGLTVAFRGHYFMMLSPSPSPSPARRITTEWGISPIDTYVT 1200
 1057 DETNINCGKPVDTGLTTLNGLTVAFRGHYFMMLSPSPSPSPARRITTEWGISPIDTYVT 1116
 1201 RCNCGKTPFPKDSQYMFNDIKDAGYPKPIFKFQGLTGOIVAAALSTAKYKMPESY 1260
 1117 RCNCGKTPFPKDSQYMFNDIKDAGYPKPIFKFQGLTGOIVAAALSTAKYKMPESY 1176
 1261 PFKRGSIQOYLYKQEPVQKCGRRPALNYPVYGEMTOVRRRRFBRALGPSQTHIRIQY 1320
 1177 PFKRGSIQOYLYKQEPVQKCGRRPALNYPVYGEMTOVRRRRFBRALGPSQTHIRIQY 1236
 1321 SPARLAIQDKGLVLAHEVYVSIIMRGLPNVVTSAISLPINRKDGVDYVAFSKOQYNNIDV 1380
 1237 SPARLAIQDKGLVLAHEVYVSIIMRGLPNVVTSAISLPINRKDGVDYVAFSKOQYNNIDV 1296
 1381 PERTARAITTRSGOTLSKWNVNC 1404

Db 1297 PSRTBARITTSRGOTLSKWNCP 13320

|||||

RESULT 7

US-07-757-022B-60

; Sequence 60, Application US/07757022B

; Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Coe, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-60

Query Match 93.2%; Score 7011; DB 4; Length 1320;

Best Local Similarity 94.0%; Pred. No. 0;

Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

Db 1 MAMTLPYLLLLSVFIQVSSQ----- 25

Qy 1 MAMTLPYLLLLSVFIQVSSQDLSSCAGCGEGSRDATCNCYNCHVCCPDF 60

Db 1 KRVCTAELISGCRGFESEFERGECDDAQCKKDKCCPDYESFCAEVNPTSPSSKAP 120

Qy 61 KRVCTAELISGCRGFESEFERGECDDAQCKKDKCCPDYESFCAEVNPTSPSSKAP 120

Db 26 -----ELSCGRCGFESFERGECDDAQCKKDKCCPDYESFCAEVNPTSPSSKAP 79

Qy 121 PPSGASQTIKSTTKRSPKPKKTKKVIIESEITEHVSSENOSSSSSSSSSSSTIW 180

Db 80 PPSGASQTIKSTTKRSPKPKKTKKVIIESEITE----- 115

Qy 181 KIKSSKNSAANRELQKLVKONKNNRKKKPTKPPVVDAGSGLDNGDKTTPDST 240

Db 116 -----VKONKNNRKKKPTKPPVVDAGSGLDNGDKTTPDST 156

Qy 241 TQHNKSVSPKITTAKPINRPSLPNSDTSKETSLTVNKEETVETKETTNNKQTSIDG 300

Db 157 TQHNKSVSPKITTAKPINRPSLPNSDTSKETSLTVNKEETVETKETTNNKQTSIDG 216

Qy 301 KEKTTSAKETOSIKTSKADLAPTSKVLAKPTPAETTKGPAITTPKEPTTPKEPAS 360

Db 217 KEKTTSAKETOSIKTSKADLAPTSKVLAKPTPAETTKGPAITTPKEPTTPKEPAS 276

Qy 361 TTPKEPTTPKSAPTTPKEPAITTPKSAPTTPKEPAITTPKEPAITTPKEPAITTPKE 420

Db 277 TTPKEPTTPKSAPTTPKEPAITTPKSAPTTPKEPAITTPKEPAITTPKEPAITTPKE 336

Qy 421 APTTKSAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 480

Db 337 APTTKSAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPK 396

Qy 481 EPAAPKAPKAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAIT 540

Db 397 EPAAPKAPKAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAIT 456

Qy 541 TTKSAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 600

Db 457 TTKSAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKE 516

Qy 601 APTAPKAPKAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAIT 660

Db 517 APTAPKAPKAPTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAIT 576

Qy 661 PEEPAITPKKAAANTKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAIT 720

Db 577 PEEPAITPKKAAANTKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAIT 636

Qy 721 APTPKKAPKELAPTTKKEPSTTSOKPAATTPKGAATTPKGAATTPKGAATTPKGA 780

Db 637 APTPKKAPKELAPTTKKEPSTTSOKPAATTPKGAATTPKGAATTPKGAATTPKGA 696

Qy 781 TAPTLKAPATTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTP 840

Db 697 TAPTLKAPATTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTPKKEPAITTP 756

Qy 841 KPAPTTPETPTTSEVSTPTTKETTHKSPEESTPELSAETPKALENSPREGVPT 900

Db 757 KPAPTTPETPTTSEVSTPTTKETTHKSPEESTPELSAETPKALENSPREGVPT 816

Qy 901 TKTPAATKPEMTTAKKTERDRTTPETTAPKMTKETAATTEKTESKITATTTQV 960

Db 817 TKTPAATKPEMTTAKKTERDRTTPETTAPKMTKETAATTEKTESKITATTTQV 876

Qy 961 TSTTTQDTPEKITTLLKTTLLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTP 1020

Db 877 TSTTTQDTPEKITTLLKTTLLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTP 936

Qy 1021 POKPTAPPKKPTSTKCKTMRVAKKPTTTPPKMTSTIMBELNPTSIAEAMLOTTTRPN 1080

Db 937 POKPTAPPKKPTSTKCKTMRVAKKPTTTPPKMTSTIMBELNPTSIAEAMLOTTTRPN 996

Qy 1081 QTPNSKLVENVPKSEEDGAGBETPHMLRPHVMPVPTDMOVLPRVPOGIIINPMLS 1140

Db 997 QTPNSKLVENVPKSEEDGAGBETPHMLRPHVMPVPTDMOVLPRVPOGIIINPMLS 1056

Qy 1141 DETNINCGKVPDGLITLRNGLVAFRGHYFWMLSPPSPSPARRITEVWGIPSDIVFT 1200

Db 1057 DETNINCGKVPDGLITLRNGLVAFRGHYFWMLSPPSPSPARRITEVWGIPSDIVFT 1116

Qy 1201 RNCCEGKTFPPKOSQYRFTNDIKDAGYKPIIFKGFGGLGQIYALSTAKYKWPSSVY 1260

Db 1117 RNCCEGKTFPPKOSQYRFTNDIKDAGYKPIIFKGFGGLGQIYALSTAKYKWPSSVY 1176

QY 1141 DETNINCKGKPYDGLTTLNGLTVAFRGHYFMWLSPEFSPSPARRITTEVMGIPSPIDVTFT 1200
DB 1048 DETNINCKGKPYDGLTTLNGLTVAFRGHYFMWLSPEFSPSPARRITTEVMGIPSPIDVTFT 1107
QY 1201 RCNCEGKTFPFKDSQYMRFTNDIKDAGYKPIFKGFGGLTGOIYAALSTAKYKMPESVY 1260
DB 1108 RCNCEGKTFPFKDSQYMRFTNDIKDAGYKPIFKGFGGLTGOIYAALSTAKYKMPESVY 1167
QY 1261 FPKKGGSIQQYITKCEPVQKCPGRPALNYPYGGMTQVRRRRFERAIGPGQTHIRIQY 1320
DB 1168 FPKKGGSIQQYITKCEPVQKCPGRPALNYPYGGMTQVRRRRFERAIGPGQTHIRIQY 1227
QY 1321 SPARLAYODKGYLHNEVKVSIIMRGLPMVVTSAISLPNIRKPDGYDYAFSKDOYNTIDV 1380
DB 1228 SPARLAYODKGYLHNEVKVSIIMRGLPMVVTSAISLPNIRKPDGYDYAFSKDOYNTIDV 1287
QY 1381 PSRTARAITTRSGQTLISKVWYNCP 1404
DB 1288 PSRTARAITTRSGQTLISKVWYNCP 1311

RESULT 9

US-07-757-022B-50
Sequence 50, Application US/07757022B

Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Geisner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 1314 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-50

Query Match 92.6%; Score 6971; DB 4; Length 1314;

Best Local Similarity 93.6%; Pred. No. 0;

Matches 1314; Conservative 0; Mismatches 0; Indels 90; Gaps 1;

QY 1 MAMKTLPIYLILLSVFIQOVSSQDLSKSCAGRGEGYSDATCNQYCNHCECPDF 60
DB 1 MAMKTLPIYLILLSVFIQOVSSQDLSKSCAGRGEGYSDATCNQYCNHCECPDF 60
QY 61 KRVCTAELSCGRCFBSFEREGCDACQCKYDKCCPDYESFCAEYHNPTSPSSKAP 120
DB 61 KRVCTAELSCGRCFBSFEREGCDACQCKYDKCCPDYESFCAEYHNPTSPSSKAP 120
QY 121 PPSASQTIKSTIKRSKPPKKTAKVISEEITTEHSVSENOSSSSSSSSSTIW 180
DB 121 PPSASQTIKSTIKRSKPPKKTAKVISEEITTEHSVSENOSSSSSSSSSTIW 180
QY 67 -----HVSVENQSSSSSSSSSSSTIW 90
DB 67 -----HVSVENQSSSSSSSSSSSTIW 90
QY 181 KIKSSKNSAANRELQKKLVKDNKKNRTKKKPPPVNDEAGSLNDGDKVTPPTST 240
DB 91 KIKSSKNSAANRELQKKLVKDNKKNRTKKKPPPVNDEAGSLNDGDKVTPPTST 150
QY 241 TQHNKVSIPKITTAKDINRPSLPPNSDTSKETSITVNNKTYETKETTNNKQTSIDG 300
DB 151 TQHNKVSIPKITTAKDINRPSLPPNSDTSKETSITVNNKTYETKETTNNKQTSIDG 210
QY 301 KEKTSKAKTOSIKTSKADLAPTSKVLAKPTPAEITTKGPAITTKPEPTTPKBPAS 360
DB 211 KEKTSKAKTOSIKTSKADLAPTSKVLAKPTPAEITTKGPAITTKPEPTTPKBPAS 270
QY 361 TTPKEPTTKSAPPTPKBPAPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420
DB 271 TTPKEPTTKSAPPTPKBPAPTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBP 330
QY 421 APTTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 480
DB 331 APTTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 390
QY 481 EPAPTAPEKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTKSAPTTKBP 540
DB 391 EPAPTAPEKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTKSAPTTKBP 450
QY 541 TTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTKBP 600
DB 451 TTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTKBP 510
QY 601 APTAPKEBAPPTPKETAAPTTPKLPTTPEKLAFTTPEKAPATTPEELAPTTPE 660
DB 511 APTAPKEBAPPTPKETAAPTTPKLPTTPEKLAFTTPEKAPATTPEELAPTTPE 570
QY 661 PEEBAPPTPKAAANTKEBAPPTPKBPAPTTPKBPAPTTPKETAAPTTKATTTLEP 720
DB 571 PEEBAPPTPKAAANTKEBAPPTPKBPAPTTPKBPAPTTPKETAAPTTKATTTLEP 630
QY 721 APTTPKAPAPKELAPTTKEPTSTSDKPAFTTGPAPTTKGPAPTTKBPAPTTPK 780
DB 631 APTTPKAPAPKELAPTTKEPTSTSDKPAFTTGPAPTTKGPAPTTKBPAPTTPK 690
QY 781 TAPTLKBPAPTTPKKAPKELAPTTKGPSTTSDDKPAFTTKEBAPTTPKBPAPTTPK 840
DB 691 TAPTLKBPAPTTPKKAPKELAPTTKGPSTTSDDKPAFTTKEBAPTTPKBPAPTTPK 750
QY 841 KPAFTTPEPTPTTSSEVSTPTTKEPTTIHKSPESTPELSAETPALENSPEEGVPT 900
DB 751 KPAFTTPEPTPTTSSEVSTPTTKEPTTIHKSPESTPELSAETPALENSPEEGVPT 810
QY 901 TKTPAATKPEMTTAKKTERDRTTETTTAAPTAKKATTEKTSKATTTATTOV 960
DB 811 TKTPAATKPEMTTAKKTERDRTTETTTAAPTAKKATTEKTSKATTTATTOV 870
QY 961 TSTTODTTPFKITTLTKTTLLAPKVTYTTKTIITTEIMNKPEETAKPRDRAVSKATTPK 1020
DB 961 TSTTODTTPFKITTLTKTTLLAPKVTYTTKTIITTEIMNKPEETAKPRDRAVSKATTPK 1020


```

Db      871 TSTTTDTPPKITLTKTLTLAPKVTTKITTTTEIMNKBETAKKORATNSKATPK 930
Qy      1021 POKPTAPKKPTSTKPKPTMVRVKPKPTPTPKMTSTMBLAPTSRIAEMLQTTTRPN 1080
Db      931 POKPTAPKKPTSTKPKPTMVRVKPKPTPTPKMTSTMBLAPTSRIAEMLQTTTRPN 990
Qy      1081 QTPNSKLVEVNPKSDEAGAGAGETPHMLLRPHVFMBEVTPDMOYLPRVNOGIIINPMLS 1140
Db      991 QTPNSKLVEVNPKSDEAGAGAGETPHMLLRPHVFMBEVTPDMOYLPRVNOGIIINPMLS 1050
Qy      1141 DETNINCKGRVNDLTLRLNGTLVAFRGHYFMMLSPSPSPSPARITBWCISPIDTTFVT 1200
Db      1051 DETNINCKGRVNDLTLRLNGTLVAFRGHYFMMLSPSPSPSPARITBWCISPIDTTFVT 1110
Qy      1201 RCNCEGTFPPKQSOQWRFTNDIKDGYPKPIFKGFGGLGQYVALSTAKYNWPEASY 1260
Db      1111 RCNCEGTFPPKQSOQWRFTNDIKDGYPKPIFKGFGGLGQYVALSTAKYNWPEASY 1170
Qy      1261 PFKRGSIOQYIKQEBVOKCPGRPALNYPVYGMTQVRRRFEBAIGPSQTHIRIQY 1320
Db      1171 PFKRGSIOQYIKQEBVOKCPGRPALNYPVYGMTQVRRRFEBAIGPSQTHIRIQY 1230
Qy      1321 SPARLAVODKGVLANEVKYSILWRGLPNTTSAISLPIRKPDGYIYAFSKQYNNIDV 1380
Db      1231 SPARLAVODKGVLANEVKYSILWRGLPNTTSAISLPIRKPDGYIYAFSKQYNNIDV 1290
Qy      1381 PSRTARAITRSGQTLKQWYNCP 1404
Db      1291 PSRTARAITRSGQTLKQWYNCP 1314

RESULT 10
US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Geener, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genetic Institute, Inc.
; STREET: 87 Cambridge Park Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geer, Luann
; REGISTRATION NUMBER: 31,822

```

```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-142

Query Match      92.4%; Score 6951.5; DB 4; Length 1313;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

Qy      1 MAMKTLPIYLILLLSVFVIQVYSSODLSSACAGCGEGYRDATCNCNDYVCOHMECCPDF 60
Db      1 MAMKTLPIYLILLLSVFVIQVYSSQ----- 25
Qy      61 KVCCTAELSCGRCESEFERGECDDACCKYDKCCPDYSEFCAEVHNPSPSSKAP 120
Db      26 -----ELSCKGRCESEFERGECDDACCKYDKCCPDYSEFCA----- 64
Qy      121 PPSGASQRTIKSTKSPKPKPKTKVYLSBEITEHVSSENOSSSSSSSSSSTIM 180
Db      65 -----EHSVSSENOSSSSSSSSSSSSTIM 89
Qy      181 KIKSSKNSAANELOKQKVKONKKNRTKKKPPKPVVDEAGSGLDNGDFVYTPDTS 240
Db      90 KIKSSKNSAANELOKQKVKONKKNRTKKKPPKPVVDEAGSGLDNGDFVYTPDTS 149
Qy      241 TQHNVSSTSPKTTTAKPINRPSLPPNSDTSKETSIVNKKETTVETKETTNNKQSTNG 300
Db      150 TQHNVSSTSPKTTTAKPINRPSLPPNSDTSKETSIVNKKETTVETKETTNNKQSTNG 209
Qy      301 KEKTSASAKTOSIEKTSANOLAPTSKVLAKPPKAEITTKGPAITTPKEPTTPKEPAS 360
Db      210 KEKTSASAKTOSIEKTSANOLAPTSKVLAKPPKAEITTKGPAITTPKEPTTPKEPAS 269
Qy      361 TTPKEPTPTTIKSAATTPKEPAITTKSAAPTPPKSAATTPKEPAITTPKEPAITTKP 420
Db      270 TTPKEPTPTTIKSAATTPKEPAITTKSAAPTPPKSAATTPKEPAITTPKEPAITTKP 329
Qy      421 APTTKSAATTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 480
Db      330 APTTKSAATTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 389
Qy      481 EPAPTAPKKAPATTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 540
Db      390 EPAPTAPKKAPATTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 449
Qy      541 TTKSAATTPKEBSPPTTKKEPAITTPKEPAITTPPKKAPATTPKEPAITTPKEPAITTKP 600
Db      450 TTKSAATTPKEBSPPTTKKEPAITTPKEPAITTPPKKAPATTPKEPAITTPKEPAITTKP 509
Qy      601 APTAKKEPAITTPKEPAITTPKEKLTTPTEKLAFTTPKEKAPATTPKEBLAFTTPBEPPT 660
Db      510 APTAKKEPAITTPKEPAITTPKEKLTTPTEKLAFTTPKEKAPATTPKEBLAFTTPBEPPT 569
Qy      661 PEEPAATPKAAAPTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 720
Db      570 PEEPAATPKAAAPTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 629
Qy      721 APTPKKAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKEPAITTPKEPAITTPK 780
Db      630 APTPKKAPKELAPTTTKEPTSTSDKAPATTPKGTATTPKEPAITTPKEPAITTPK 689
Qy      781 TAPTLKKEPAITTPKAPKELAPTTTGPSTSDKAPATTPKGTATTPKEPAITTPK 840
Db      690 TAPTLKKEPAITTPKAPKELAPTTTGPSTSDKAPATTPKGTATTPKEPAITTPK 749
Qy      841 KPAPTPEPPTTSEVSTPTTKEPTTIHKSDESTBELSAEPTPKALENSPKRGPVT 900

```

Db 750 KPAPTPTPTPTTSEVSTPTTKEPTTIHKSPTDSTBELSAPFKALENSPKRGVPT 809
QY 901 TKTPATYPEMTTAKDXTTERDOLTTPBTATAAPMTKETATTEKTESKIATTVQV 960
Db 810 TKTPATPEMTTAKDXTTERDOLTTPBTATAAPMTKETATTEKTESKIATTVQV 869
QY 961 TSTTODTTPFKITTLTKTTLLAPKYTTTKITITTEINKKPEETAKPDORANSATTPK 1020
Db 870 TSTTODTTPFKITTLTKTTLLAPKYTTTKITITTEINKKPEETAKPDORANSATTPK 929
QY 1021 POKPTAKKPTSTKPKTPMVRKPKTTPPRKMTSTMPBELNPTSRITAEAMLOTTTSPN 1080
Db 930 POKPTAKKPKPTSTKPKTPMVRKPKTTPPRKMTSTMPBELNPTSRITAEAMLOTTTSPN 989
QY 1081 QTPNSKLVEVNPKSDAGABSETPHMLLRPHVMEVETPDMDYLPRVNOGIIINPMLS 1140
Db 990 QTPNSKLVEVNPKSDAGABSETPHMLLRPHVMEVETPDMDYLPRVNOGIIINPMLS 1049
QY 1141 DETNLCNKPVDGLTTLNGLTVAFRGHYFWMLSPSPSPARRITETVWGI PSPIDTVFT 1200
Db 1050 DETNLCNKPVDGLTTLNGLTVAFRGHYFWMLSPSPSPARRITETVWGI PSPIDTVFT 1109
QY 1201 RNCCEGKTFEFDSDQYWFETNDIKDAGYPKPIFKGFGJTGQIYAALSTAKYKMPESVY 1260
Db 1110 RNCCEGKTFEFDSDQYWFETNDIKDAGYPKPIFKGFGJTGQIYAALSTAKYKMPESVY 1169
QY 1261 FPKGGSIOQYIYYKQEPVQKCPRRPALNYPYGGMTQVRRRRFRAIGPSQHTTIRIQY 1320
Db 1170 FPKGGSIOQYIYYKQEPVQKCPRRPALNYPYGGMTQVRRRRFRAIGPSQHTTIRIQY 1229
QY 1321 SPARLAYODKGVLANEVSILMRGLPVVYVSALSIPNIRKPDGYDYAFSDQYVNIIDV 1380
Db 1230 SPARLAYODKGVLANEVSILMRGLPVVYVSALSIPNIRKPDGYDYAFSDQYVNIIDV 1289
QY 1381 PSRTARAITTTRSGOTLSKVMYNCP 1404
Db 1290 PSRTARAITTTRSGOTLSKVMYNCP 1313

RESULT 11
US-07-757-022B-44
Sequence 44, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Coert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: G1 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-44

Query Match 89.4%; Score 6727; DB 4; Length 1270;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAMKTLPIYLLILLISVPIQOVSSODLSSCAGRGEGSRNATCNCDVNCQHWECPDF 60
Db 1 MAMKTLPIYLLILLISVPIQOVSSQ----- 25
QY 61 KRVCIAELSCGRGCFESFEREGRECDCAQCKYDKCPDYESFCAEVANPTSPSSKAP 120
Db 26 -----ELSCGRGCFESFEREGRECDCAQCKYDKCPDYESFCAE----- 65
QY 121 PPSGASQTIKSTKRSXPKNKKTKKYIESEETIEHSVSENQSSSSSSSSSSTIW 180
Db 66 ----- 65
QY 181 KIKSSKNSAANRELQKKLVKONKKNRTKKKPKPPVVDAGSGLDNGDFKTTPTDST 240
Db 66 -----VDNKKKNRTKKKPKPPVVDAGSGLDNGDFKTTPTDST 106
QY 241 TQHNKVSPTKITTAKKINPRLPSPNSDTSKETSLTVNKETVETKETTNNKQSTDG 300
Db 107 TQHNKVSPTKITTAKKINPRLPSPNSDTSKETSLTVNKETVETKETTNNKQSTDG 166
QY 301 KEKTTSAKETOSIKTSAKDIAPTSKVLAKPTPAETTTGPAITTPKEPTTPKBPAS 360
Db 167 KEKTTSAKETOSIKTSAKDIAPTSKVLAKPTPAETTTGPAITTPKEPTTPKBPAS 226
QY 361 TTPKEPTTPKSAPTTPKBPAPTTTSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 420
Db 227 TTPKEPTTPKSAPTTPKBPAPTTTSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 286
QY 421 APTTTSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 480
Db 287 APTTTSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 346
QY 481 EPAPTAPEKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 540
Db 347 EPAPTAPEKAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 406
QY 541 TTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 600
Db 407 TTKSAPPTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 466
QY 601 APTAPKEPAPTTPKETAPTTPKULTPTPEKLAAPTTPEKAPTTPEELAPTTPEEPTPT 660
Db 467 APTAPKEPAPTTPKETAPTTPKULTPTPEKLAAPTTPEKAPTTPEELAPTTPEEPTPT 526
QY 661 PEEBAPTTPKAAADNTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 720
Db 527 PEEBAPTTPKAAADNTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 586

QY 601 APTAPKEBAPTTPEKATPTPKLPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTT 660
DB 601 APTAPKEBAPTTPEKATPTTPEKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPTTT 660
QY 661 PEEBAPTTPKAAAPTPKEBAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 720
DB 661 PEEBAPTTPKAAAPTPKEBAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 720
QY 721 APTTPKAPKEAPTTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 780
DB 721 APTTPKAPKEAPTTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 780
QY 781 TAPTTKEBAPTTPEKPAKEAPTTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEK 840
DB 781 TAPTTKEBAPTTPEKPAKEAPTTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEK 840
QY 841 KPAPTTPPTTSEVSTPTTKEPTTTHKSPESTPELSAAPTPKALENSPKPGVPT 900
DB 841 KPAPTTPPTTSEVSTPTTKEPTTTHKSPESTPELSAAPTPKALENSPKPGVPT 900
QY 901 TKTATATPEMTTAKDITERRDLATTPETTTAAAPKMTKEATTTTEKTESKITATTTQV 960
DB 901 TKTATATPEMTTAKDITERRDLATTPETTTAAAPKMTKEATTTTEKTESKITATTTQV 960
QY 961 TSTTTODTTPPKITTLTKTTTAPKVTTKITTTTEINMKPEETAKPKORATNSKATTPK 1020
DB 961 TSTTTODTTPPKITTLTKTTTAPKVTTKITTTTEINMKPEETAKPKORATNSKATTPK 1020
QY 1021 POKPTAPKPKPTSTKPKTPMRVRKPTTTPRKYTSIMPELNPSTIAEAMLOTTTTPN 1080
DB 1021 POKPTAPKPKPTSTKPKTPMRVRKPTTTPRKYTSIMPELNPSTIAEAMLOTTTTPN 1080
QY 1081 QTPNSKLVNVPKSDAGAGETPHMLLRHVMPETPTPMDYIPRPNGOIIINPMLS 1140
DB 1081 QTPNSKLVNVPKSDAGAGETPHMLLRHVMPETPTPMDYIPRPNGOIIINPMLS 1140

RESULT 13

US-07-757-022B-58
Sequence 58, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-58

Query Match 73.2%; Score 5510.5; DB 4; Length 1049;
Best Local Similarity 92.0%; Pred. No. 0; Mismatches 0; Indels 91; Gaps 2;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAMKTLPIYLLLLSVFVIQVSSQDSSCAGRCGEGYSDATCNCYNCHYMECCPDF 60
DB 1 MAMKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVCATLSECKGRFPESFERGECDCDAQCKYKCCPDYHSFCAEYVHNPTSPSSKAP 120
DB 26 -----ELSCGRCRFBESFERGECDCDAQCKYKCCPDYHSFCA----- 64
QY 121 PPGASQTIKSTTRSRKPPKTKKVISEETIEHSVENESSSSSSSSSSSTIw 180
DB 65 -----EHSVSVENESSSSSSSSSSSTIw 89
QY 181 KIKSSNSAANRELQKLVKDNKNTKKKPTPKPVVDAGSLDNGDFKVTTPDTST 240
DB 90 KIKSSNSAANRELQKLVKDNKNTKKKPTPKPVVDAGSLDNGDFKVTTPDTST 149
QY 241 TQHNKVSIPKITTAKAINPSPSLPNSDTSKETSILVNKETTVEKETTNTNQSTDG 300
DB 150 TQHNKVSIPKITTAKAINPSPSLPNSDTSKETSILVNKETTVEKETTNTNQSTDG 209
QY 301 KEKTTSAKETOSIKTSAKDLAPTSKYLAKPTPAEITTTGPAITTPKEBPTTPKEPAS 360
DB 210 KEKTTSAKETOSIKTSAKDLAPTSKYLAKPTPAEITTTGPAITTPKEBPTTPKEPAS 269
QY 361 TTPKEPTPTTIKSAPTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTTKEB 420
DB 270 TTPKEPTPTTIKSAPTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTTKEB 329
QY 421 APTTTKSAPTTPKEBAPTTTPKKAAPTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTTKEB 480
DB 330 APTTTKSAPTTPKEBAPTTTPKKAAPTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTTKEB 389
QY 481 EPAPTAPKKAAPTTPKEBAPTTTPKKAAPTTPKESPTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAP 540
DB 390 EPAPTAPKKAAPTTPKEBAPTTTPKKAAPTTPKESPTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAP 449
QY 541 TTKSAPTTPKEBAPTTTKEBAPTTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKA 600
DB 450 TTKSAPTTPKEBAPTTTKEBAPTTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKA 509
QY 601 APTAPKEBAPTTPEKATPTTPEKLTPTTPEKLAPTTPEKPAPTTPEEELAPTTPEEPTTT 660
DB 510 APTAPKEBAPTTPEKATPTTPEKLTPTTPEKLAPTTPEKPAPTTPEEELAPTTPEEPTTT 569
QY 661 PEEBAPTTPKAAAPTPKEBAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 720
DB 570 PEEBAPTTPKAAAPTPKEBAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 629
QY 721 APTTPKAPKEAPTTTKEPTSTSDKAPPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEK 780

Db 630 APTPKKAPKAPKELAPTTKEPTSTSDKPAPTTKGTAPTPKAPPTPKAPPTPKG 689
 Qy 781 TAPTLKEPAPPTPKKAPKAPKELAPTTKGPTSTSDKPAPTTKETAPTPKAPPTPK 840
 Db 690 TAPTLKEPAPPTPKKAPKAPKELAPTTKGPTSTSDKPAPTTKETAPTPKAPPTPK 749
 Qy 841 KPAPTTPEPTPTSEVSTPTTKEPTTIHKSPPDESIPBELSAEPTPALENSPEKGVPT 900
 Db 750 KPAPTTPEPTPTSEVSTPTTKEPTTIHKSPPDESIPBELSAEPTPALENSPEKGVPT 809
 Qy 901 TKTPTAATKPEMTTAAKDKTERDLRTPEPTTAAAPKATKETAATTEKTESKITATTTQV 960
 Db 810 TKTPTAATKPEMTTAAKDKTERDLRTPEPTTAAAPKATKETAATTEKTESKITATTTQV 869
 Qy 961 TSTTTODTTPKTTTLTKTTTLAPKVTTKTTTETIMNKPEETAKKDAATSKATTPK 1020
 Db 870 TSTTTODTTPKTTTLTKTTTLAPKVTTKTTTETIMNKPEETAKKDAATSKATTPK 929
 Qy 1021 POKPTTAPKKTSTKPKTPMPVRKPKTTPTPKMTSTMPBLNPTSRIAEAMLTQTTTPN 1080
 Db 930 POKPTTAPKKTSTKPKTPMPVRKPKTTPTPKMTSTMPBLNPTSRIAEAMLTQTTTPN 989
 Qy 1081 QTFNSKLVENVPKSEBAGAEGETPHMLLRPHVFMPEVTPDMDYLPKVPNOGIIINPMLS 1140
 Db 990 QTFNSKLVENVPKSEBAGAEGETPHMLLRPHVFMPEVTPDMDYLPKVPNOGIIINPMLS 1049
 RESULT 14
 US-07-757-022B-84
 / Sequence 84, Application US/07757022B
 / Patent No. 6433142
 / GENERAL INFORMATION:
 / APPLICANT: Geener, Thomas C.
 / APPLICANT: Clark, Stephen C.
 / APPLICANT: Turner, Katherine
 / APPLICANT: Hewick, Rodney M.
 / TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 / NUMBER OF SEQUENCES: 143
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Genetics Institute, Inc.
 / STREET: 87 Cambridgepark Drive
 / CITY: Cambridge
 / STATE: Massachusetts
 / COUNTRY: U.S.A.
 / ZIP: 02140
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/07/757,022B
 / FILING DATE: 19910910
 / CLASSIFICATION: 530
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/643,502
 / FILING DATE: 18-JAN-1991
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/546,114
 / FILING DATE: 29-JUN-1990
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/457,196
 / FILING DATE: 29-DEC-1989
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/390,901
 / FILING DATE: 08-AUG-1989
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Geert, Luann
 / REGISTRATION NUMBER: 31,822
 / REFERENCE/DOCKET NUMBER: GI 5190
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (617) 876-1170
 / TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 84:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1022 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-84
 Query Match 72.1%; Score 5425.5; DB 4; Length 1022;
 Best Local Similarity 91.5%; Pred. No. 0;
 Matches 1020; Conservative 0; Mismatches 2; Indels 93; Gaps 2;
 Db 26 DISSCAGRCGEYSRDATCCOYHMECCPDFRVCYTAELSCGRCFESRERED 85
 1 DISSCAGRCGEYSRDATCCOYHMECCPDFRVC--ELSCGRCFESRERED 58
 Qy 86 CDAQCKKYKCCPDYSEFCAEVHNPTSPSSKKAAPPSSAQTIKSTKSPKPKKKT 145
 Db 59 CDAQCKKYKCCPDYSEFCAEV----- 80
 Qy 146 KAYISEEITTEHSVSENQESSSSSSSSSTIWKIKSSKQSANRELQKLVKONK 205
 Db 81 -----AVADNKK 87
 Qy 206 NRTKKKPPKPPVVDAGSGLDNGDPKVTTPDSTTHNKKVSTSPKITTAKPINRPSLP 265
 Db 88 NRTKKKPPKPPVVDAGSGLDNGDPKVTTPDSTTHNKKVSTSPKITTAKPINRPSLP 147
 Qy 266 PMSDTSKETSLTVNKKETVEYKETTNNKQSTDGKEKTSKAKETQSIKTSKADLAPTS 325
 148 PMSDTSKETSLTVNKKETVEYKETTNNKQSTDGKEKTSKAKETQSIKTSKADLAPTS 207
 Qy 326 KYLAKPTPAEETTTGPAITTPKEPTPTPKBPASTTPKEPTTIKSAPTTPKEBAPT 385
 Db 208 KYLAKPTPAEETTTGPAITTPKEPTPTPKBPASTTPKEPTTIKSAPTTPKEBAPT 267
 Qy 386 TKSAPTPEBAPTTPPKBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 445
 Db 268 TKSAPTPEBAPTTPPKBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 327
 Qy 446 TTPKEBAPTTPKEPTTPPKBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPK 505
 Db 328 TTPKEBAPTTPKEPTTPPKBAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPK 387
 Qy 506 PAPTTPKSPPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 565
 Db 388 PAPTTPKSPPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 447
 Qy 566 KEPAPTTPKAPPTTPKEBAPTTPKEBAPTTPKKAAPTAPKEBAPTTPKEBAPTTP 625
 Db 448 KEPAPTTPKAPPTTPKEBAPTTPKEBAPTTPKKAAPTAPKEBAPTTPKEBAPTTP 507
 Qy 626 PTTPEKLAATTPPEKAPPTTPEELATTPPEEPTTPPEBAPTTPKAAANTPKBAPTTP 685
 Db 508 PTTPEKLAATTPPEKAPPTTPEELATTPPEEPTTPPEBAPTTPKAAANTPKBAPTTP 567
 Qy 686 KEPAPTTPKEBAPTTPKETAPTTPGYNAPTTLKEBAPTTPKAPKAPKELAPTTKEPTST 745
 Db 568 KEPAPTTPKEBAPTTPKETAPTTPGYNAPTTLKEBAPTTPKAPKAPKELAPTTKEPTST 627
 Qy 746 SDKPAPTTPKGAAPTTPKEBAPTTPKBPATTPKGAAPTTLKBPAPTTPKAPKAPKELAPT 805
 Db 628 SDKPAPTTPKGAAPTTPKEBAPTTPKBPATTPKGAAPTTLKBPAPTTPKAPKAPKELAPT 687
 Qy 806 TTGPTSTTSDKPAPTTKETAATTPPKBAPTTPKKAAPTTPKETPTPTTSEVSTPTTKE 865
 Db 688 TTGPTSTTSDKPAPTTKETAATTPPKBAPTTPKKAAPTTPKETPTPTTSEVSTPTTKE 747
 Qy 866 PTTIHKSPDESIPBELSAEPTPALENSPEKGVPTTKTPAATKPEMTTAAKDKTERDLR 925
 Db 748 PTTIHKSPDESIPBELSAEPTPALENSPEKGVPTTKTPAATKPEMTTAAKDKTERDLR 807
 Qy 926 TTPETTTAAPKATKETAATTEKTESKITATTTQVTSITTTODTTPFKITTLTTLAPRV 985

DB 808 TTPETTAAPKWTKEATATTEKTESKITATTTQVTSSTTQDTTFFKITLLKTLLAKV 867
QY 966 TTTKTTTITTEIMNKEPEETAKPKDAINSKATTPKQPKTKPKKPTSTKPKKPMRYRK 1045
DB 868 TTTKTTTITTEIMNKEPEETAKPKDAINSKATTPKQPKTKPKKPTSTKPKKPMRYRK 927
QY 1046 PKTPTPKMTSTWELNPTSRIAEAMLOTTTRPNOTPKSLVEVNPSEADGAGETP 1105
DB 928 PKTPTPKMTSTWELNPTSRIAEAMLOTTTRPNOTPKSLVEVNPSEADGAGETP 987
QY 1106 HMLRPHVFMPEVTEDMDYLPRVNOGIIINPMLS 1140
DB 988 HMLRPHVFMPEVTEDMDYLPRVNOGIIINPMLS 1022
RESULT 15
US-07-757-022B-74
Sequence 74, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridge Park Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-74

Query Match 70.9%; Score 5335; DB 4; Length 1038;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1038; Conservative 0; Mismatches 0; Indels 366; Gaps 3;

QY 1 MAMKTLPIYLLLSVFVIOOVSSODLSSCAGRCGEGSRDATGCDYNQGHMECCPDF 60
DB 1 MAMKTLPIYLLLSVFVIOOVSSODLSSCAGRCGEGSRDATGCDYNQGHMECCPDF 25
QY 61 KRVTAEISCKGRGFESEFEREGECDAQCCKYKDCPDYBSCAEVHNPSPSSKAP 120
DB 26 -----ELSCGRGFESEREGECDAQCCKYKDCPDYBSCAEVHNPSPSSKAP 65
QY 121 PPSGASQTIKSTYRSKPPNKKTKKYIESEITTEHSVSENDESSSSSSSSSTIW 180
DB 66 -----
QY 181 KIKSSKSAANRELQKKLVKONKKNRTKKKPKPVVDEAGSGLDNGDFKVTTPDTST 240
DB 66 -----VDONKKNRTKKKPKPVVDEAGSGLDNGDFKVTTPDTST 106
QY 241 TQHNKVSPTPKITAKPINRPSLPPNSDTSKETSLSLVNKEETVETKETTNNKQSTDG 300
DB 107 TQHNKVSPTPKITAKPINRPSLPPNSDTSKETSLSLVNKEETVETKETTNNKQSTDG 166
QY 301 KEKTTSAKETOSIKTSKADLAFTSKYLAKTPKAEITTKGPAITTPKEPTTPKBPAS 360
DB 167 KEKTTSAKETOSIKTSKADLAFTSKYLAKTPKAEITTKGPAITTPKEPTTPKBPAS 226
QY 361 TTPKEPTTPKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 420
DB 227 TTPKEPTTPKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 286
QY 421 APTTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 480
DB 287 APTTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 346
QY 481 EPAPTAKKKAPPTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 540
DB 347 EPAPTAKKKAPPTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 406
QY 541 TTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 600
DB 407 TTKSAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 466
QY 601 APTAPKBPAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 660
DB 467 APTAPKBPAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 526
QY 661 PEBAPTTPKKAAANTKEBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 720
DB 527 PEBAPTTPKKAAANTKEBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 586
QY 721 APTPKKAPAPKELAPTTTKBPSTSDKPAPTTPKGAFTTPKBPAPTTTPKBPAPTTTPKBP 780
DB 587 APTPKKAPAPKELAPTTTKBPSTSDKPAPTTPKGAFTTPKBPAPTTTPKBPAPTTTPKBP 646
QY 781 TAPTLKEBPAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 840
DB 647 TAPTLKEBPAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 706
QY 841 KPAPTTTETPTTSEVSTPTTKETPHKSPBSTPELSAETPKALNSPREPVPT 900
DB 707 KPAPTTTETPTTSEVSTPTTKETPHKSPBSTPELSAETPKALNSPREPVPT 766
QY 901 TKTPAATKPEMTTAKKTERDRLTTPETTTAPKMTKETAATTEKTESKITATTTQV 960
DB 767 TKTPAATKPEMTTAKKTERDRLTTPETTTAPKMTKETAATTEKTESKITATTTQV 826
QY 961 TSITTDOTPEKITLLKTLLAPKVTTTKKTIITTEIMNKEPEETAKPKDAINSKATTPK 1020
DB 827 TSITTDOTPEKITLLKTLLAPKVTTTKKTIITTEIMNKEPEETAKPKDAINSKATTPK 886
QY 1021 POKPTKAPKPTSTKKEKTMRYVAKKPTTTPKMTSTWELNPTSRIAEAMLOTTTRPN 1080
DB 887 POKPTKAPKPTSTKKEKTMRYVAKKPTTTPKMTSTWELNPTSRIAEAMLOTTTRPN 946

```

Qy 1081 QTPNSKLVENVPKSEBAGAGETPHMLLRPHVFMPEVTEDMDYLPRVPMOGIINPMLS 1140
Db 947 QTPNSKLVENVPKSEBAGAGETPHMLLRPHVFMPEVTEDMDYLPRVPMOGIINPMLS 1006
Qy 1141 DETNIGNGKPVGDLTLRNGTLVAFRGHYFMMLSPFSPSPARRITEWGIISPIDTVFT 1200
Db 1007 ----- 1006
Qy 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIVALSTAKYKWPESVY 1260
Db 1007 ----- 1006
Qy 1261 PFKRGGSIQOYIYKQBPVQKCPGRPALNYPYGEVTQVRRRFRERAIGPSQTHIRIQY 1320
Db 1007 ----- 1006
Qy 1321 SPARLAYQDKVLAHEVKVSIILMRGLPNVTSAISLEPIRKPDGYDYAFESKQYNNIDY 1380
Db 1007 -----DOYNNIDY 1014
Qy 1381 PSRTARAITTRSGGTLISKVWYNCP 1404
Db 1015 PSRTARAITTRSGGTLISKVWYNCP 1038

```

Search completed: December 8, 2003, 09:38:09
 Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:19:41 ; Search time 54 Seconds
(without alignments)
4126.891 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526

Sequence: 1 MAMKTLPIYLILLLSVFVIQ.....AAATTRSGQTLSKVMYVNC 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	13	MSF precursor. Sy
2	7526	100.0	1404	22	Human megakaryocyte
3	7526	100.0	1404	22	Human megakaryocyte
4	7523	100.0	1415	22	Novel human secret
5	6950	92.3	1299	22	Human EST encoded
6	3484	46.3	902	22	Human MSF-derived
7	2929	38.9	851	22	Human testes-deriv
8	2920	38.8	546	22	Human testes-deriv
9	2850.5	37.9	538	23	3' cartilage super

10	2757	36.6	513	22	ABU53254	Human testes-deriv
11	2197	29.2	452	16	AA880041	Human megakaryocyte
12	1707.5	22.7	472	22	AA860569	Bovine MSF ortholog
13	1545	20.5	292	23	AA011261	Human HAP0 polypep
14	1188	15.8	5179	22	AA824516	C8939 predicted am
15	1188	15.8	5179	24	ABP55456	Human colon tumour
16	1097	14.6	214	22	ABU53255	Human testes-deriv
17	1012.5	13.5	717	22	ABU53144	Human testes-deriv
18	981	13.0	763	21	AA838842	Arabidopsis thaliana
19	968.5	12.9	188	23	AA018833	5' cartilage super
20	950	12.6	1664	19	AA443106	C. thermocellum O1
21	946.5	12.6	1049	22	AB861364	Drosophila melanog
22	927.5	12.3	778	22	ABU53143	Human testes-deriv
23	927	12.3	770	22	ABU53141	Human testes-deriv
24	902	12.0	1795	22	AB869806	Drosophila melanog
25	900.5	12.0	717	22	ABU53145	Human testes-deriv
26	900.5	12.0	717	22	ABU53146	Human testes-deriv
27	900.5	12.0	717	22	ABU53147	Human testes-deriv
28	900.5	12.0	717	22	ABU53148	Human testes-deriv
29	900.5	12.0	717	22	ABU53149	Human testes-deriv
30	900.5	12.0	717	22	ABU53150	Human testes-deriv
31	900.5	12.0	717	22	ABU53151	Human testes-deriv
32	890	11.8	745	22	ABU53142	Human testes-deriv
33	889	11.8	692	22	ABU53155	Human testes-deriv
34	848.5	11.3	745	22	ABU53154	Human testes-deriv
35	818.5	10.9	695	22	ABU53152	Human testes-deriv
36	805	10.7	1325	22	AB849735	Human liver peptid
37	805	10.7	1325	22	AB829725	Peptide #2403 enco
38	805	10.7	1325	22	AB834897	Peptide #2403 enco
39	805	10.7	1325	22	AB820314	Protein #2313 enco
40	805	10.7	1325	22	AA855707	Human brain expres
41	805	10.7	1325	22	AA866085	Human testes-deriv
42	805	10.7	1325	22	AA803645	Human testes-deriv
43	805	10.7	1325	22	AB837612	Human liver peptid
44	782.5	10.4	212	23	AB860403	Human peptide enco
45	776	10.3	395	22	ABU53160	Drosophila melanog

ALIGNMENTS

RESULT 1					
AA826049					
ID	AA826049	standard:	Protein:	1404	AA.
AC	AA826049;				
XX					
DT	25-MAR-2003	(updated)			
DT	02-FEB-1993	(first entry)			
XX					
DE	MSF precursor.				
XX					
KW	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;				
KW	stability; proteolytic cleavage; adhesion; alternative splicing.				
OS	Synthetic.				
XX					
EH	Key	Location/Qualifiers			
FT	Region	1..26			
FT	Region	/label= Exon_I			
FT	Region	26..67			
FT	Region	/label= Exon_II			
FT	Region	67..107			
FT	Region	/label= Exon_III			
FT	Region	107..157			
FT	Region	/label= Exon_IV			
FT	Region	157..200			
FT	Region	/label= Exon_V			
FT	Region	200..1141			
FT	Region	/label= Exon_VI			
FT	Region	1411..1166			
FT	Region	/label= Exon_VII			
FT	Region	1166..1212			

FT /label= Exon_VIII
FT 1213..1266
FT /label= Exon_IX
FT 1266..1331
FT /label= Exon_X
FT 1331..1373
FT /label= Exon_XI
FT 1373..1404
FT /label= Exon_XII
FN MO9213075-A1.
PD 06-AUG-1992.
XX 17-JAN-1992; 92MO-US00433.
XX 18-JAN-1991; 91US-0643502.
PR 10-SEP-1991; 91US-0757022.
XX (GENY) GENETICS INST INC.
PA
XX Clark SC, Geaner TG, Hewick RM, Jacobs K, Turner K;
PI
XX MPI; 1992-284660/34.
DR N-PSDB; AAQ27223.
XX
XX New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.
XX
XX Claim 1, 2 and 3; Fig 1, 87bp; English.
XX
XX The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine proceeding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNAs each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, ie. one or more of these exons may contain
CC sequences which direct proteolytic cleavage, adhesion, organisation of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occurring and non-naturally occurring MSFs may be characterised by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.
CC (Updated on 25-MAR-2003 to correct FN field.)
XX
SQ Sequence 1404 AA;
Query Match 100.0%; Score 7526; DB 13; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAMTLPYLLLLSVFVIOQVSSQDSSSCAGRCGEGSRDATGNCNDYCNHYMECCPDF 60
DB 1 MAMTLPYLLLLSVFVIOQVSSQDSSSCAGRCGEGSRDATGNCNDYCNHYMECCPDF 60
QY 61 KRVTABISCKGRCPESFERGBCDCAQCKKXDKCPCDYSPCAEVNPTSPSSKAP 120
DB 61 KRVTABISCKGRCPESFERGBCDCAQCKKXDKCPCDYSPCAEVNPTSPSSKAP 120

QY 121 PPSGASOTIKSTTRSPKPKKKTKKYIESEETIEHSVENOESSSSSSSSSTTW 180
DB 121 PPSGASOTIKSTTRSPKPKKKTKKYIESEETIEHSVENOESSSSSSSSSTTW 180
QY 181 KIKSSKNSAANRELQKLVKONKONRTKKKTPKPVVDAGSGLDNGDKVTPDTST 240
DB 181 KIKSSKNSAANRELQKLVKONKONRTKKKTPKPVVDAGSGLDNGDKVTPDTST 240
QY 241 TQHNKVSSTPKITAKPLNPPPSLPNSDTSKETSILVNKETTETTTNNQSTDG 300
DB 241 TQHNKVSSTPKITAKPLNPPPSLPNSDTSKETSILVNKETTETTTNNQSTDG 300
QY 301 KEKTTSAKETOSIKETSAKDLAPTSKYLAKPTPAETTTKGPALTTPKEPTTPKEBAS 360
DB 301 KEKTTSAKETOSIKETSAKDLAPTSKYLAKPTPAETTTKGPALTTPKEPTTPKEBAS 360
QY 361 TTPKEPTTPTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPT 420
DB 361 TTPKEPTTPTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPT 420
QY 421 APTTTKSAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 480
DB 421 APTTTKSAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 480
QY 481 EPAPTAPKKAAPTTPKEBAPTTPKEBAPTTPKESPTTPKEBAPTTPKSAPTTPKEBAPT 540
DB 481 EPAPTAPKKAAPTTPKEBAPTTPKEBAPTTPKESPTTPKEBAPTTPKSAPTTPKEBAPT 540
QY 541 TTKSAPTTPKEBAPTTPKEBAPTTPKKAAPTTPKKAAPTTPKEBAPTTPKKAAPT 600
DB 541 TTKSAPTTPKEBAPTTPKEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 600
QY 601 APTAPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 660
DB 601 APTAPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 660
QY 661 PEBAPAPTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 720
DB 661 PEBAPAPTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 720
QY 721 APTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 780
DB 721 APTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 780
QY 781 TAPPTLKBPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 840
DB 781 TAPPTLKBPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 840
QY 841 KPAATTEPTPTTSEVSTPTTKEPTTIHKSPESTPELSAETPALENSPEKPVPT 900
DB 841 KPAATTEPTPTTSEVSTPTTKEPTTIHKSPESTPELSAETPALENSPEKPVPT 900
QY 901 TKTPAATKPEMTTAKDTERDLITTEFTTAAPKMTKETATTTKTESKITATTTQV 960
DB 901 TKTPAATKPEMTTAKDTERDLITTEFTTAAPKMTKETATTTKTESKITATTTQV 960
QY 961 TSTTTQDTPPKITLTKTTLAPKVTYTTKTIITTEIMNKPBEBAKXDAATNSKATTPK 1020
DB 961 TSTTTQDTPPKITLTKTTLAPKVTYTTKTIITTEIMNKPBEBAKXDAATNSKATTPK 1020
QY 1021 POKPTKAPKPTSPKSKPTPRVAKKPTTPTPRKMTSTMBELNPTSLIAEMLQTTTTPN 1080
DB 1021 POKPTKAPKPTSPKSKPTPRVAKKPTTPTPRKMTSTMBELNPTSLIAEMLQTTTTPN 1080
QY 1081 QTPNSKLVEVNPKSSEDAGAGETPHMLLRPHVMPVETPDMDYLPRVNPQGIINPMLS 1140
DB 1081 QTPNSKLVEVNPKSSEDAGAGETPHMLLRPHVMPVETPDMDYLPRVNPQGIINPMLS 1140
QY 1141 DETNINCKGKPYDGLTTLRNGTLVAFRGHYFMWLSPEPSPBARITTEWGIPTDVT 1200
DB 1141 DETNINCKGKPYDGLTTLRNGTLVAFRGHYFMWLSPEPSPBARITTEWGIPTDVT 1200
QY 1201 RCNCEGKTFPFKDSQYVRFTNDIKDAGYPRIFKFGGLNGQIVAAALSTAKYKMPESVY 1260

```

DB 1201 RCNCEGTEFFKDSQYWRFTNDIKDGYKPIFKGFGGLTGQVVALSTAKVNMPESTY 1260
QY 1261 PFRFGSIOQYTYKQBPVQKCGRRPALVYVGEVTOVRRRRFEALIPSGHTIRIOY 1320
DB 1261 PFRFGSIOQYTYKQBPVQKCGRRPALVYVGEVTOVRRRRFEALIPSGHTIRIOY 1320
QY 1321 SPARLAYQDKVLAHNEVKVSLMRGLPNVYTSALPNIRKPDGYDYAFSKDQYNNIDV 1380
DB 1321 SPARLAYQDKVLAHNEVKVSLMRGLPNVYTSALPNIRKPDGYDYAFSKDQYNNIDV 1380
QY 1381 PSTARAITTRSGQTLISKVYNCP 1404
DB 1381 PSTARAITTRSGQTLISKVYNCP 1404

RESULT 2
AAB60568
ID AAB60568 standard; Protein; 1404 AA.
AC AAB60568;
XX 27-APR-2001 (first entry)
DT Human megakaryocyte stimulating factor (MSF, CACP).
DE Human megakaryocyte stimulating factor (MSF, CACP).
XX Human; CACP protein; campodactyly-arthropathy-coxa vara-pericarditis;
KM MSF; megakaryocyte stimulating factor; synovial lubricant;
KM chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
KM antiarthritic.
XX Homo sapiens.
OS MO200107068-A1.
PN 01-FEB-2001.
PD 21-JUL-2000; 2000MO-US20002.
PF 23-JUL-1999; 99US-0145328.
PR 19-JUL-2000; 2000US-0145328.
XX (UYCA-) UNIV CASE WESTERN RESERVE.
PA Warman ML;
PI WPI; 2001-182721/18.
DR New composition comprising the campodactyly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints.
XX Example 1; Page -; 34p; English.
XX The invention relates to a method of treating osteoarthritis via the
XX administration of a composition comprising the campodactyly-arthropathy-
XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX The composition may further comprise a local anesthetic. The composition
XX of the invention may be administered via intra-articular or intravenous
XX injection. The human CACP protein is identified in the invention as
XX being megakaryocyte stimulating factor (MSF). The gene encoding
XX CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
XX arthropathy-coxa vara-pericarditis, in which patients have synovial
XX hyperplasia without evidence of inflammation. CACP protein (MSF)
XX acts as a synovium lubricant, and can be used to lubricate tissue and
XX joints in the treatment of osteoarthritis. The composition may be
XX applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX loss of range of movement or joint damage). The present sequence
XX represents human megakaryocyte stimulating factor (MSF, CACP protein).
XX Note: This sequence is not given in its entirety in figure 4 of the
XX specification, although a Genbank accession number was given. This
XX sequence was therefore obtained from Genbank (U70316)

```

```

XX SQ Sequence 1404 AA;
Query Match 100.0%; Score 7526; DB 22; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKLPYLLLLSVFIQOVSSODLSSCAGRCGEYSRDACTCNDNCCQHYMECCPDF 60
DB 1 MAMKLPYLLLLSVFIQOVSSODLSSCAGRCGEYSRDACTCNDNCCQHYMECCPDF 60
QY 61 KRVCTAELSCGRCESEFERGECDDAQCCKKYDCCPDYESFCAEVNPTSPSSKAP 120
DB 61 KRVCTAELSCGRCESEFERGECDDAQCCKKYDCCPDYESFCAEVNPTSPSSKAP 120
QY 121 PPSGASQITIKSTTKSPKPPNKKTKKVISEBEITEHSVSENOSSSSSSSSSTTW 180
DB 121 PPSGASQITIKSTTKSPKPPNKKTKKVISEBEITEHSVSENOSSSSSSSSSTTW 180
QY 181 KIKSSKNSAANRELQKLVKNDKKNRTKKKPTPKPVPVDEAGSLDNGDFKVTTPDST 240
DB 181 KIKSSKNSAANRELQKLVKNDKKNRTKKKPTPKPVPVDEAGSLDNGDFKVTTPDST 240
QY 241 TQHNKYSTSPKITTAKPINRPSLPPNSDTSEKTSLVNKKETVYTKETTTNNKQSTDG 300
DB 241 TQHNKYSTSPKITTAKPINRPSLPPNSDTSEKTSLVNKKETVYTKETTTNNKQSTDG 300
QY 301 KEKTTSAKETOSIEKTSADLAFTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360
DB 301 KEKTTSAKETOSIEKTSADLAFTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360
QY 361 TTPKEPTPTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 420
DB 361 TTPKEPTPTTKSAPTTKEBPATTTKSAPTTKEBPATTTKSAPTTKEBPATTTKEP 420
QY 421 APTTTSAPTTKKEBPATTTKKEBPATTTKKEBPATTTKKEBPATTTKKEBPATTTK 480
DB 421 APTTTSAPTTKKEBPATTTKKEBPATTTKKEBPATTTKKEBPATTTKKEBPATTTK 480
QY 481 EPAPTAAPKAPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 540
DB 481 EPAPTAAPKAPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 540
QY 541 TTKSAPTTKKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 600
DB 541 TTKSAPTTKKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 600
QY 601 APTAPKEBPATTTKETAATTTKELJPTTPEKLAAPTTPKEKAPATTPESEIAPTTP 660
DB 601 APTAPKEBPATTTKETAATTTKELJPTTPEKLAAPTTPKEKAPATTPESEIAPTTP 660
QY 661 PEEBPATTTKAAAPNTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 720
DB 661 PEEBPATTTKAAAPNTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 720
QY 721 APTTPKKAPELAAPTTPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 780
DB 721 APTTPKKAPELAAPTTPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 780
QY 781 TAPTLKEBPATTTKKEBPATTTKGPSTTSDDKAPATTPEKATATTTPKEBPATTT 840
DB 781 TAPTLKEBPATTTKKEBPATTTKGPSTTSDDKAPATTPEKATATTTPKEBPATTT 840
QY 841 KPAPTTPPTPTTSEVSTPTTKKEPTTIHKSPESESTBELSABPTPKALNSPKGVP 900
DB 841 KPAPTTPPTPTTSEVSTPTTKKEPTTIHKSPESESTBELSABPTPKALNSPKGVP 900
QY 901 TKTPAATKEMTTAKDRTTERTTAAAPKMTKETATTEKTESKITATTTQV 960
DB 901 TKTPAATKEMTTAKDRTTERTTAAAPKMTKETATTEKTESKITATTTQV 960
QY 961 TSTTQDTPPKITTLKTTTLAPKVTTKKTTITTEINNKKEETAKEPDRATNSKATTPK 1020
DB 961 TSTTQDTPPKITTLKTTTLAPKVTTKKTTITTEINNKKEETAKEPDRATNSKATTPK 1020

```

DB 961 TSTTQDPTPKITTLTKTTTLTAFLKVTYTKITITTEIMNKKPEETAKPRORATNSKATTPK 1020
QY 1021 POKPTKAPKKTSTSTKPKTKMTRVRKPKTTPPRKKTSTMPBELNPTSRITAEAMLQTTTTPN 1080
DB 1021 POKPTKAPKKTSTSTKPKTKMTRVRKPKTTPPRKKTSTMPBELNPTSRITAEAMLQTTTTPN 1080
QY 1081 QTPNSKIVENVPKSDAGAEGETHMLLRPHVFEVETPDMDYLPVYNOGIIINPMLS 1140
DB 1081 QTPNSKIVENVPKSDAGAEGETHMLLRPHVFEVETPDMDYLPVYNOGIIINPMLS 1140
QY 1141 DETNINCKPVDGLTTLNGLTVAFRGHYFMMWSPSPSPSPARRITEVWGIPSPIDVT 1200
DB 1141 DETNINCKPVDGLTTLNGLTVAFRGHYFMMWSPSPSPSPARRITEVWGIPSPIDVT 1200
QY 1201 RONEGKTFPPFDSQYMFRTNDIKAGYPKPIFKFGSLTQIYVAALSTAKYKMPESVY 1260
DB 1201 RONEGKTFPPFDSQYMFRTNDIKAGYPKPIFKFGSLTQIYVAALSTAKYKMPESVY 1260
QY 1261 FFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFRAIGPOTHTIRIQY 1320
DB 1261 FFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFRAIGPOTHTIRIQY 1320
QY 1321 SPARLAYODKGVLANEVKVSILMRGLPNVVTSAISLPIRKPDGYDYAFAFSKQYINIDV 1380
DB 1321 SPARLAYODKGVLANEVKVSILMRGLPNVVTSAISLPIRKPDGYDYAFAFSKQYINIDV 1380
QY 1381 PSRTARATITRSQGLTSKVMYVNC 1404
DB 1381 PSRTARATITRSQGLTSKVMYVNC 1404

RESULT 3

AAB29773
ID AAB29773 standard; Procein; 1404 AA.

AC AAB29773;

DT 28-FEB-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX Human MSF; megakaryocyte stimulating factor; tribonecin;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KM osteoarthritis; tribonecin; tissue adhesion inhibition;
KM friction coefficient reduction; gene therapy; antiarthritic;
KM osteopathic.

XX Homo sapiens.

XX OS

XX MO200064930-A2.

XX PD 02-NOV-2000.

XX PF 24-APR-2000; 2000MO-US10953.

XX PR 23-APR-1999; 99US-0298970.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX JAY GD;

XX MPI: 2001-024673/03.

XX DR N-PSDB; AAC81498.

XX Novel tribonecin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety

XX Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonecin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonecin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76

CC repeats of a motif having at least 50% identity to the sequence KSPAPPT
CC (AAB29773). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonecin; a biocompatible composition comprising a
CC human tribonecin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonecin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonecin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonecin may be used in gene therapy. The present sequence represents
CC human MSF.

CC Sequence 1404 AA;

CC Query Match 100.0%; Score 7526; DB 22; Length 1404;

CC Best Local Similarity 100.0%; Pred. No. 0;

CC Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKTLPIYLILLSVFIQVSSODLSSCGRCGEGYRATGNCNCHWECPCDF 60

DB 1 MAMKTLPIYLILLSVFIQVSSODLSSCGRCGEGYRATGNCNCHWECPCDF 60

QY 61 KRVCIAELSCGRCFESFERGECDCDQCKYKDCPDYESFCAEVHNPTSPSSKAP 120

DB 61 KRVCIAELSCGRCFESFERGECDCDQCKYKDCPDYESFCAEVHNPTSPSSKAP 120

QY 121 PPSGASQTIKSTKRSKPKPKKKTKYVISEEITEHSVSENQSSSSSSSSSTIW 180

DB 121 PPSGASQTIKSTKRSKPKPKKKTKYVISEEITEHSVSENQSSSSSSSSSTIW 180

QY 181 KIKSSKNSAANRELQKLVKNDKNTKKKPPKPVVVBAGSLGNDGPKVTPPTST 240

DB 181 KIKSSKNSAANRELQKLVKNDKNTKKKPPKPVVVBAGSLGNDGPKVTPPTST 240

QY 241 TQHNKVSPTKITTAKINPSPSLPNSDTSKETSILVNNKETTVEKTTNNKQTSIDG 300

DB 241 TQHNKVSPTKITTAKINPSPSLPNSDTSKETSILVNNKETTVEKTTNNKQTSIDG 300

QY 301 KEKTTSAKETOSIKTSADKDIAPTSKVLAKPTPAETTTKGPALTTPKEPTTPKEPAS 360

DB 301 KEKTTSAKETOSIKTSADKDIAPTSKVLAKPTPAETTTKGPALTTPKEPTTPKEPAS 360

QY 361 TTPKEPTTTKSAPTTPKEAPPTTTSAPTTPKEAPPTTTPKEAPPTTTPKEAP 420

DB 361 TTPKEPTTTKSAPTTPKEAPPTTTSAPTTPKEAPPTTTPKEAPPTTTPKEAP 420

QY 421 APITTKSAPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPK 480

DB 421 APITTKSAPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPK 480

QY 481 EPAPTAPEKAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 540

DB 481 EPAPTAPEKAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 540

QY 541 TTKSAPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 600

DB 541 TTKSAPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 600

QY 601 APAPKAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 660

DB 601 APAPKAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 660

QY 661 PEEPAPTTPKAAADNTEKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 720

DB 661 PEEPAPTTPKAAADNTEKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 720

```

QY 721 APTTPKKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTPKKAPATTTPKEBAPTTPKG 780
DB 721 APTTPKKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTTPKEBAPTTPKEBAPTTPKG 780
QY 781 TAPTTTKEBAPTTPPKKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTTPKEBAPTTPK 840
DB 781 TAPTTTKEBAPTTPPKKAPKELAPTTTKEPTSTSDKPAPTTPKGTAPTTTPKEBAPTTPK 840
QY 841 KPAPTTPETPTTSEVSTPTTKEBPTTHKSPDESPBELSAPPTPALENSKEBQVPT 900
DB 841 KPAPTTPETPTTSEVSTPTTKEBPTTHKSPDESPBELSAPPTPALENSKEBQVPT 900
QY 901 TKTPTAPKPEMTTADKTERDLRTTPTTAAAPKMTKEATTTTEKTESKITATTTQV 960
DB 901 TKTPTAPKPEMTTADKTERDLRTTPTTAAAPKMTKEATTTTEKTESKITATTTQV 960
QY 961 TSTTTTQDTPPKKTTTLKTTTLAPKVTTTKTTTTEIMNKPEBETAKKDBATNSKATTPK 1020
DB 961 TSTTTTQDTPPKKTTTLKTTTLAPKVTTTKTTTTEIMNKPEBETAKKDBATNSKATTPK 1020
QY 1021 POKPTAPKPKSTTKKPKTMPRVAKPTTTPRKMSTMBELNPTSRIAEAMLTQTTTRV 1080
DB 1021 POKPTAPKPKSTTKKPKTMPRVAKPTTTPRKMSTMBELNPTSRIAEAMLTQTTTRV 1080
QY 1081 QTENSKLVEVNPKESEDAGAEGETPHMLLRPHVFMVEVTPDMOYLPRVPMQIINPMLS 1140
DB 1081 QTENSKLVEVNPKESEDAGAEGETPHMLLRPHVFMVEVTPDMOYLPRVPMQIINPMLS 1140
QY 1141 DEFNICKGKVPDGLTTLRNGTLVAFRGHYFMMLSPSPSPSAPARRITEWGI PPSIDTVEFT 1200
DB 1141 DEFNICKGKVPDGLTTLRNGTLVAFRGHYFMMLSPSPSPSAPARRITEWGI PPSIDTVEFT 1200
QY 1201 RNCBEGTTPFKSOQVWRFTNDIKDAGYPRKIFRGREGTLGQVVAALSTKYNWPEASY 1260
DB 1201 RNCBEGTTPFKSOQVWRFTNDIKDAGYPRKIFRGREGTLGQVVAALSTKYNWPEASY 1260
QY 1261 FFKRGGSIQOYIYKQEBVQKCPGRRRPALNPVYGEHTQVRRRFEBAIGSQTHTTRIQY 1320
DB 1261 FFKRGGSIQOYIYKQEBVQKCPGRRRPALNPVYGEHTQVRRRFEBAIGSQTHTTRIQY 1320
QY 1321 SPARLAYQDKGLAHNEVKVSI LMRGLPNVVTSAISLPNIRKPDGYDYAFSSKOYYNIDV 1380
DB 1321 SPARLAYQDKGLAHNEVKVSI LMRGLPNVVTSAISLPNIRKPDGYDYAFSSKOYYNIDV 1380
QY 1381 PSRTABAITRRSGOTLSKMYNCP 1404
DB 1381 PSRTABAITRRSGOTLSKMYNCP 1404
RESULT 4
AAUJ2262
ID AAUJ2262 standard; Protein; 1415 AA.
XX
AC AAUJ2262;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2753.
XX
KM Human; vaccination; gene therapy; nutritional supplement;
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001MO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.

```

```

XX (HYSE-) HYSBQ INC.
PA Tang YT, Liu C, Dmanac RT;
PI WPI; 2001-611725/70.
XX
DR Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
PT Claim 20; Page 573; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAUJ29510-AAUJ3304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 1415 AA;

```

```

Query Match 100.0%; Score 7523; DB 22; Length 1415;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1403; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMKTLPIYLLLLSVFVYIQVSSODLSSCAGRCGEYSRDATCNDYVCOHYMECCPFP 60
DB 12 MAMKTLPIYLLLLSVFVYIQVSSODLSSCAGRCGEYSRDATCNDYVCOHYMECCPFP 71
QY 61 KEVCTAELSCRCRCESEFERGECDDAQCKKYDKCPCDYSFCAEVHNPSPSSKAP 120
DB 72 KAVCTAELSCRCRCESEFERGECDDAQCKKYDKCPCDYSFCAEVHNPSPSSKAP 131
QY 121 PPSGASQITTKSTTKSPKPPNKKTKVIESBEITBEHSVSNQSSSSSSSSSTTW 180
DB 132 PPSGASQITTKSTTKSPKPPNKKTKVIESBEITBEHSVSNQSSSSSSSSSTTW 191
QY 181 KIKSSKNSAANRELQKKLVKONKKNRTKKKPTPKPPVVDAGSLDNGDFVYTPPDST 240
DB 192 KIKSSKNSAANRELQKKLVKONKKNRTKKKPTPKPPVVDAGSLDNGDFVYTPPDST 251
QY 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETS LTVNKETTVETKETTITNKQSTDG 300
DB 252 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETS LTVNKETTVETKETTITNKQSTDG 311
QY 301 KEKTTSAKETOSIEKTSAXDLAPTSKVLAKPTPKAETTTKGALTTPKBPTTPKEPAS 360
DB 312 KEKTTSAKETOSIEKTSAXDLAPTSKVLAKPTPKAETTTKGALTTPKBPTTPKEPAS 371
QY 361 TTPKPEPTPTTISAATTTKEBAPTTPKSAPTTPKBPATTTKEBAPTTPKBPATTTKEP 420
DB 372 TTPKPEPTPTTISAATTTKEBAPTTPKSAPTTPKBPATTTKEBAPTTPKBPATTTKEP 431
QY 431 APTTTKSAAPTTPKEBAPTTPKBPATTTKEBAPTTPKBPATTTTPKEBAPTTPK 480
DB 432 APTTTKSAAPTTPKEBAPTTPKBPATTTKEBAPTTPKBPATTTTPKEBAPTTPK 491
QY 481 BPAPTAAPKAPATTTPKEBAPTTPKBPATTTKEBAPTTPKBPATTTTPKEBAPT 540
DB 492 BPAPTAAPKAPATTTPKEBAPTTPKBPATTTKEBAPTTPKBPATTTTPKEBAPT 551
QY 541 TTKSAPTTPKEBAPTTPKEBAPTTPKBPATTTTPKEBAPTTPKBPATTTTPK 600

```

Dd		552	TTKSAPTTPKESSPITTTKEBPATTTKEBPAPTTPEKPAPTTPKEBAPTTTCKP	611
Oy		601	APTAKBPAPTTPKETAFTPPKKLTPTTEBKLAFTTPEKAPTTPEBLAFTTPEEBPTT	660
Dd		612	APTAKBPAPTTPKETAFTPPKKLTPTTEBKLAFTTPEKAPTTPEBLAFTTPEEBPTT	671
Oy		661	PEEPAPTTPKAAAVNTPKEBPATTTKEBPAPTTPEKAPTTPEKTAFTTPKGTAFTTLKBP	720
Dd		672	PEEPAPTTPKAAAVNTPKEBPATTTKEBPAPTTPEKAPTTPEKTAFTTPKGTAFTTLKBP	731
Oy		721	APTPEKKAPBKLAFPTTKEPTSTSDKAPFTTPGTAFTTPEKAPTTPKETAPTTPEK	780
Dd		732	APTPEKKAPBKLAFPTTKEPTSTSDKAPFTTPGTAFTTPEKAPTTPEKAPTTPEK	791
Oy		781	TAPTTLKBPAPTTTPKKPAPEKELAPTTTGKPTSTSDKAPTTPKETAFTTPEKAPTTPK	840
Dd		792	TAPTTLKBPAPTTTPKKPAPEKELAPTTTGKPTSTSDKAPTTPKETAFTTPEKAPTTPK	851
Oy		841	KCAPPTPEPTPTTSBSVSTPTTKEPTTHKSDDSETPLSAEPTPKALENSPKBGVPT	900
Dd		852	KCAPPTPEPTPTTSBSVSTPTTKEPTTHKSDDSETPLSAEPTPKALENSPKBGVPT	911
Oy		901	TKTPAPATPEMTTAKDKTERDILTTPBTPTTAAAPMTKETATTTKEKTESKITLATTQV	960
Dd		912	TKTPAPATPEMTTAKDKTERDILTTPBTPTTAAAPMTKETATTTKEKTESKITLATTQV	971
Oy		961	TSTTTODTTPPFKITLLKTTTTLAPKYVTTKITITTEINMKPEETAKPKORANSKATTPK	1020
Dd		972	TSTTTODTTPPFKITLLKTTTTLAPKYVTTKITITTEINMKPEETAKPKORANSKATTPK	1031
Oy		1021	PQKPTPAKPKPTSTYCKPKTMPVRKXKPTTPTPRKNTSTMBELNPTRSIAEAMLQTTRBN	1080
Dd		1032	PQKPTPAKPKPTSTYCKPKTMPVRKXKPTTPTPRKNTSTMBELNPTRSIAEAMLQTTRBN	1091
Oy		1081	QTPNSKLVENVNKSSBDAGABGETPHMLRPVFMPEVLPDMDYLPVFNQGIITNPMIS	1144
Dd		1092	QTPNSKLVENVNKSSBDAGABGETPHMLRPVFMPEVLPDMDYLPVFNQGIITNPMIS	1155
Oy		1141	DETNICNGKPVVDGLTTLNRGTLVAFRGHYFMMILSPFSPPSAPARITEWGISPIDVYET	1200
Dd		1152	DETNICNGKPVVDGLTTLNRGTLVAFRGHYFMMILSPFSPPSAPARITEWGISPIDVYET	1211
Oy		1201	RNCCEGKTFEFKDSQYMRFTNDIKDAGYPKPIFKPGGLTGOIVAAISTAKYNMPSVY	1266
Dd		1212	RNCCEGKTFEFKDSQYMRFTNDIKDAGYPKPIFKPGGLTGOIVAAISTAKYNMPSVY	1277
Oy		1261	FFKRGGSIQQYYIKQEPVOQCGRRPALNYPYEGMTQVRRRFRERAIQPSQTHIRIQY	1322
Dd		1272	FFKRGGSIQQYYIKQEPVOQCGRRPALNYPYEGMTQVRRRFRERAIQPSQTHIRIQY	1333
Oy		1321	SPARALAYODKGVLANHEVKVSIIMRGLENVTSAISLPNIRKDGVDYAFASXDQYNNIDV	1386
Dd		1332	SPARALAYODKGVLANHEVKVSIIMRGLENVTSAISLPNIRKDGVDYAFASXDQYNNIDV	1397
Oy		1381	PSRTARAITTRSGQTLISKVWYNC P 1404	
Dd		1392	PSRTARAITTRSGQTLISKVWYNC P 1415	
 RESULT 5 AAM24322 ID ID AAM24322 standard; Protein; 1299 AA.				
AC		AAM24322;		
XX				
DT		12-OCT-2001 (first entry)		
XX				
DE		Human EST encoded protein SEQ ID NO: 1847.		
XX				
KM		Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;		
KM		tomato; monkey; dog; sea urchin; expressed sequence tag; EST;		
KW		diagnostics; forensic test; gene mapping; genetic disorder;		

Query Match	Best Local Similarity	92.3%; Score 6950; DB 22; Length 1299;	Matches 1295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	bio diversity; gene therapy; nutrition.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200154477-A2.		
XX	XX		
PD	02-AUG-2001.		
XX	XX		
PF	25-JAN-2001; 2001WO-US02687.		
XX	XX		
PR	25-JAN-2000; 2000US-0491404.		
PR	17-JUL-2000; 2000US-0617746.		
PR	03-AUG-2000; 2000US-0631451.		
PR	15-SEP-2000; 2000US-0663870.		
XX	XX		
PA	(HYSE-) HYSEQ INC.		
XX	XX		
PI	Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Aesund V;		
PI	Cao Y, Dermanac RA, Zhang J, Wetman T;		
XX	XX		
DR	WPI; 2001-476164/51.		
XX	XX		
DR	N-PEDB; AAH98981.		
XX	XX		
PT	Isolated polypeptide for treatment of diseases, diagnostics, raising		
PT	antibodies and research use -		
XX	XX		
PS	Claim 20; Page 1198-1201; 1275bp; English.		
XX	XX		
CC	The present invention provides the protein and coding sequences of novel		
CC	proteins from a variety of organisms, including human, dog, cat, horse,		
CC	cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea		
CC	urchin and tomato. These were derived from expressed sequence tags (ESTs)		
CC	from the organism of interest. They can be used in diagnostics,		
CC	forensics, gene mapping, identification of mutations, to assess		
CC	biodiversity and for nutritional purposes. The present sequence is a		
CC	protein of the invention.		
XX	XX		
SQ	Sequence 1299 AA;		
Query Match	92.3%; Score 6950; DB 22; Length 1299;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 1295; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MAMKTLPIYLILLLSVFVIQGVSSQDSSCAGRCGEGRSDATONCDNYCHMECCPDF	60	
DB	1 MAMKTLPIYLILLLSVFVIQGVSSQDSSCAGRCGEGRSDATONCDNYCHMECCPDF	60	
QY	61 KRVCIAELISCGRCFESFERGEREDCDACCKVYKCCPDYVSFCAEVHNPTSPSSKAP	120	
DB	61 KRVCIAELISCGRCFESFERGEREDCDACCKVYKCCPDYVSFCAEVHNPTSPSSKAP	120	
QY	121 PPSGASQTIKSTTKRSPKPKKKTKVIESEETIEHSVSENOESSSSSSSSSTIWM	180	
DB	121 PPSGASQTIKSTTKRSPKPKKKTKVIESEETIEHSVSENOESSSSSSSSSTIWM	180	
QY	181 KIKSSKNSAANRELQKLVKNDKNNKRTKKKPTPKPVVDASGGLDNGDFKVTPTDST	240	
DB	181 KIKSSKNSAANRELQKLVKNDKNNKRTKKKPTPKPVVDASGGLDNGDFKVTPTDST	240	
QY	241 TOHNVKSTISPKITTAAPINPPSLPNSDSIKESLSLTNKKTYVETKTTTNNKQTSIDG	300	
DB	241 TOHNVKSTISPKITTAAPINPPSLPNSDSIKESLSLTNKKTYVETKTTTNNKQTSIDG	300	
QY	301 KEKTTSAETOSIEKTSKADLAPTSKVLAKPTPKAETTKGEPALTTPKPTPKEPAS	360	
DB	301 KEKTTSAETOSIEKTSKADLAPTSKVLAKPTPKAETTKGEPALTTPKPTPKEPAS	360	
QY	361 TTPKEPTPTTIIKSAPTPKKEBAPTITTSAPITPKBAPTITTKBAPTITTKBAPTITTKB	420	
DB	361 TTPKEPTPTTIIKSAPTPKKEBAPTITTSAPITPKBAPTITTKBAPTITTKBAPTITTKB	420	
QY	421 APTTTKSAPTTPKEBAPTTPKKBAPTTPKEBAPTTPKAPTTPPKBAPTTPKAPTTPK	480	

```

Db 421 APTTKSAPTTKEBPAPTTPKKBPAPTTKEBPAPTTPTTKEBPAPTTKEBPAPTTPK 480
Oy 481 BPAPTPAPKBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTTSAPTTKEBPAP 540
Db 481 BPAPTPAPKBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTTSAPTTKEBPAPTT 540
Oy 541 TTKSAPTTKEBPAPTTKEBPAPTTPKKBPAPTTKEBPAPTTKEBPAPTTTKEBPAPTTTKEBP 600
Db 541 TTKSAPTTKEBPAPTTKEBPAPTTKEBPAPTTPKKBPAPTTKEBPAPTTTKEBPAPTTTKEBP 600
Oy 601 APTAKBPAPTTPKKBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTTKEBP 660
Db 601 APTAKBPAPTTPKKBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTTKEBP 660
Oy 661 PEEBPAPTTPKKBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTTKEBP 720
Db 661 PEEBPAPTTPKKBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTKEBPAPTTTKEBP 720
Oy 721 APTTKKBPAPKELAPTTTKEBPSTSDKBPAPTTPKGIAVTTKEBPAPTTKEBPAPTTKEBPAPTT 780
Db 721 APTTKKBPAPKELAPTTTKEBPSTSDKBPAPTTPKGIAVTTKEBPAPTTKEBPAPTTKEBPAPTT 780
Oy 781 TAPTTKEBPAPTTPKKBPAPKELAPTTTKEBPSTSDKBPAPTTPKGIAVTTKEBPAPTTTKEBPAP 840
Db 781 TAPTTKEBPAPTTPKKBPAPKELAPTTTKEBPSTSDKBPAPTTPKGIAVTTKEBPAPTTTKEBPAP 840
Oy 841 KPAPTTPEPPTTSEVSTPTTKEPTTHKSPDSTPELSAPPTPKLENSPKBPVPT 900
Db 841 KPAPTTPEPPTTSEVSTPTTKEPTTHKSPDSTPELSAPPTPKLENSPKBPVPT 900
Oy 901 TKTPAATKPEMTTAKDTEHDLRTTETETAAPKMTKEATTTTEKTESKITATTTQV 960
Db 901 TKTPAATKPEMTTAKDTEHDLRTTETETAAPKMTKEATTTTEKTESKITATTTQV 960
Oy 961 TSTTQDTTTPFKITTLTKTTTAPKVTYTTKTIITTEINMKPEETAKPKORATNSKATTPK 1020
Db 961 TSTTQDTTTPFKITTLTKTTTAPKVTYTTKTIITTEINMKPEETAKPKORATNSKATTPK 1020
Oy 1021 POKPPTAKKPKSTTKKPTMPVNRKPTTPTPKMTSTMPBELNPTSRILAEMLQTTTTPN 1080
Db 1021 POKPPTAKKPKSTTKKPTMPVNRKPTTPTPKMTSTMPBELNPTSRILAEMLQTTTTPN 1080
Oy 1081 QTPNSKLEVPNPKSDAGAGETPHMLLRPHVFMPEVTPDMDYTPRVNOCIIINPMLS 1140
Db 1081 QTPNSKLEVPNPKSDAGAGETPHMLLRPHVFMPEVTPDMDYTPRVNOCIIINPMLS 1140
Oy 1141 DETNINCNGKPVNGLTTLNGLVAFRGHYFMMLSPSPSPPARITTEVWGIPSPIDTVFT 1200
Db 1141 DETNINCNGKPVNGLTTLNGLVAFRGHYFMMLSPSPSPPARITTEVWGIPSPIDTVFT 1200
Oy 1201 RCNCGKTPFPKDSQYWFNTNDIKDAGYPKPIFKGFGULTGQIVAAALSTAKTKWMBESVY 1260
Db 1201 RCNCGKTPFPKDSQYWFNTNDIKDAGYPKPIFKGFGULTGQIVAAALSTAKTKWMBESVY 1260
Oy 1261 FFKRGSGIQOYIYKQEPVQKCGRRPALNYPVYGE 1295
Db 1261 FFKRGSGIQOYIYKQEPVQKCGRRPALNYPVYGE 1295

```

```

KW osteopathic.
OS Homo sapiens.
FN WO200064930-A2.
PD 02-NOV-2000.
PF 24-APR-2000; 2000WO-US10953.
PR 23-APR-1999; 99US-0238970.
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
PI Jay GD:
DR WPI; 2001-024673/03.
PT Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety
XX Disclousure; Fig 1; 47p; English.
PS
CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KEPAPTT
CC (AAB29774). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonectin; a biocompatible composition comprising a
CC human tribonectin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC or trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonectin may be used in gene therapy. The present sequence represents
CC a substantial portion of a human MSF-derived tribonectin.
XX
SQ Sequence 902 AA;
Query Match 46.3%; Score 3484; DB 22; Length 902;
Best Local Similarity 74.6%; Pred. No. 1.9e-179;
Matches 794; Conservative 28; Mismatches 80; Indels 162; Gaps 66;
Oy 200 VKDNKKKRTKKKPTPKBPVNDGAGSLDNGDFKVTTPDTSTTQHNKYSTSKITTAKEPN 259
Db 1 VKDNKKKRTKKKPTPKBPVNDGAGSLDNGDFKVTTPDTSTTQHNKYSTSKITTAKEPN 60
Oy 260 PRPSLPNPSDTSKSTSLTANKETTVETKETTNTTKQSTDSKERTTSAKTSQSIKTSK 319
Db 61 PRPSLPNPSDTSKSTSLTANKETTVETKETTNTTKQSTDSKERTTSAKTSQSIKTSK 120
Oy 320 DLAPTSKVLAKPTKAEFTTGGPALTTTPKEPTTTPKEBPASTTTPKEBPPTTIKAPTTPK 379
Db 121 DLAPTSKVLAKPTKAEFTTGGPALTTTPKEBPASTTTPKEBPPTTIKAPTTPK 172
Oy 380 EPAPTTTSAPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 439
Db 173 EPAPTTTSAPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 228
Oy 440 PKKAPPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBP 499
Db 229 -KEBPAPTT-KEBPAPTT-KEBPAPTT-KEBPAPTTTKEBPAPTTT-KEBPAPTT-KEBPAPTT-KEBP 281
Oy 500 PTTKEBPAPTTTKEBPSTTTPKEBPAPTTTSAPTTKEBPAPTTTSAPTTKEBPSTTTTKE 559

```



```
Db 282 PTT-KEBAP--TTKEBAPTT--KEBAPTTKEBAP--TTKEBAPTTKEBAPTT--KEBAP--TTKE 335
Qy 560 PAPPTKEBAPPTTKKAPPTTKBAPPTTKBAPPTTKKAPAPTBAPBAPPTTKBAPT 619
Db 336 PAPPT--KEBAPPT--KEBAPPT--KEBAP--TTKEBAPPT--KEBAPPT--KEBAPT 388
Qy 620 TPKLTPTPEKAPPTPEKAPPTPEBAPPTPEBAPPTPEBAPPTPEBAPPTPEBAPPTPE 679
Db 389 T-KEBAPTTKEP--APTKE--PAPTT--KEBAPTT--KEBAPPT--KEBAPPT--KE 440
Qy 680 PAPPTKEBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTK 739
Db 441 PAPPT--KEBAPPT--KEBAPPT--KEBAPPT--KEBAPPT--KEBAPPT--KEBAP--TTK 488
Qy 740 EPSTSDKAPPTPKGAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAP 799
Db 489 EPAPPTKE--PAPPT--KEBAPPT--KEBAPPT--KEBAPPT--KEBAPPT--KEBAPT 541
Qy 800 KELAPPTTKGPTSTSDKAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTK 859
Db 542 ----TTKEBAPPTKE--PAPTT--KEBAPPT--KEBAPPT--KEBAPPTKE--PAPTTKEPA-- 589
Qy 860 PTTKEBPTTHKSPDESTPELSAPPTPKALNSPKBGPPTTKTAPATKPEMTTAKDXT 919
Db 590 -PTTKEBAPTTKEBAPTT--KEBAP--TTKEP--APTKEBAPPTKEBAPPTKE-- 636
Qy 920 TERDLRTPTPTTAPAPTKTATTTTEKTESKITATTTQVTSSTTQDTPFKITLTKTT 979
Db 637 ----PAPT-----TKBAPPT-----KEBAPPTKEBAPPTKEBAP----- 667
Qy 980 TLAAPTITTKITTTTEIMNKEETAKPKORATNSKATTPPKQKPTAPKAPKPTSTKKPKT 1039
Db 668 ----TTKEBAPPTKEP--APT--TKBP-- 686
Qy 1040 MPVRKPKPTPTPKMTSTMBELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPXSBDAGC 1099
Db 687 ----APTPTPKMTSTMBELNPTSRIAEAML--TTTRPNQTPNSKLVEVNPXSBDAGC 739
Qy 1100 AEGTPEMLLRPHVPMRBYTPDMYLRVNPQGIINPMLSDENINCKGKAVDGLTTLRN 1159
Db 740 AEGTPEMLLRPHVPMRBYTPDMYLRVNPQGIINPMLSDENINCKGKAVDGLTTLRN 799
Qy 1160 GTLVAFGHYFWMMLSPSPSPAPARITEVWGISPIDVFTRCNCEGKTFEFDKSOYWRP 1219
Db 800 GTLVAFGHYFWMMLSPSPSPAPARITEVWGISPIDVFTRCNCEGKTFEFDKSOYWRP 859
Qy 1220 TNDIKDAGYPRKIFKGFGLTGOIUALSTAKYNWESVYFFK 1263
Db 860 TNDIKDAGYPRKIFKGFGLTGOIUALSTA--YKNWESVYFFK 902

RESULT 7
ABUS3253
ID ABUS3253 standard; Protein; 551 AA.
XX
AC ABUS3253;
XX
XX 14-APR-2003 (first entry)
XX
OS Homo sapiens.
XX
XX MO200112659-A2.
XX
XX 22-FEB-2001.
XX
XX 18-AUG-2000; 2000MO-IB01496.
XX
XX 18-AUG-1999; 99US-0149499.
PR
```

```
PR 28-SEP-1999; 99US-0156503.
XX
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
DR WPI; 2001-327840/34.
XX
XX Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies -
XX
XX Example III; Page 892-893; 1095pp; English.
XX
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention.
XX
SQ Sequence 551 AA;
XX
Query Match 38.9%; Score 2929; DB 22; Length 551;
Best Local Similarity 100.0%; Pred. No. 8,3e-150; Indels 0; Gaps 0;
Matches 551; Conservative 0; Mismatches 0;

Qy 597 TKKAPAPKEBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTK 656
Db 1 TKKAPAPKEBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTKBAPPTTK 60
Qy 657 TPTPEBAPPTPKAABNTPKEBAPPTKEBAPPTKEBAPPTPKGAPTT 716
Db 61 TPTPEBAPPTPKAABNTPKEBAPPTKEBAPPTKEBAPPTPKGAPTT 120
Qy 717 LKEBAPPTPKKAPKELAPPTTKBPTSTSDKAPPTPKGAPTTKEBAPPTKEBAPT 776
Db 121 LKEBAPPTPKKAPKELAPPTTKBPTSTSDKAPPTPKGAPTTKEBAPPTKEBAPT 180
Qy 777 TPKGAPTTKEBAPPTPKKAPKELAPPTTKGPTSTSDKAPPTPKGAPTTKEBAPT 836
Db 181 TPKGAPTTKEBAPPTPKKAPKELAPPTTKGPTSTSDKAPPTPKGAPTTKEBAPT 240
Qy 837 TTPKAPPTPEPTPTTSVSTPTTKEPTTHKSPDESTPELSAPPTPKALNSPKB 896
Db 241 TTPKAPPTPEPTPTTSVSTPTTKEPTTHKSPDESTPELSAPPTPKALNSPKB 300
Qy 897 GVPTTKTPATKPEMTTAKDXTTERDLRTPTPTTAAPKMTKEBATTTEKTESKITAT 956
Db 301 GVPTTKTPATKPEMTTAKDXTTERDLRTPTPTTAAPKMTKEBATTTEKTESKITAT 360
Qy 957 TTVQVSTTTQDTPFKITLTKTTTLAPKVTYTKKITTITTEIMNKEETAKPKORATNSKA 1016
Db 361 TTVQVSTTTQDTPFKITLTKTTTLAPKVTYTKKITTITTEIMNKEETAKPKORATNSKA 420
Qy 1017 TTPKQKPTAPKPKPTSTKPKTPRVRKXTTPTPKMTSTMBELNPTSRIAEAMLQTT 1076
Db 421 TTPKQKPTAPKPKPTSTKPKTPRVRKXTTPTPKMTSTMBELNPTSRIAEAMLQTT 480
Qy 1077 TRPNQTPNSKLVEVNPXSBDAGAGBTPMLLRPHVFMERBYTPDMYLRVNPQGIIN 1136
Db 481 TRPNQTPNSKLVEVNPXSBDAGAGBTPMLLRPHVFMERBYTPDMYLRVNPQGIIN 540
Qy 1137 PMLSDENINCN 1147
Db 541 PMLSDENINCN 551

RESULT 8
ABUS3252
```


SQ Sequence 538 AA:
 Query Match 37.9%; Score 2850.5; DB 23; Length 538;
 Best Local Similarity 99.8%; Pred. No. 1.4e-145;
 Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY PTHHSDESEPELSAETPKALNSPKPEGVPTTKTPAATKPMPTTAKDTERDUR 925
 DB 1 PTHHSDESEPELSAETPKALNSPKPEGVPTTKTPAATKPMPTTAKDTERDUR 60
 QY 926 TTPETTAAPKPKETATTEKTESKITATTTQVTSSTTQDTPFKITLTKTTTAPKV 985
 DB 61 TTPETTAAPKPKETATTEKTESKITATTTQVTSSTTQDTPFKITLTK-TTAPKV 119
 QY 986 TTTKTTTITTEIMNDEETAKPKDRAATSKATTPPKQPKTAPKPKPTSTKPKPMRVRK 1045
 DB 120 TTTKTTTITTEIMNDEETAKPKDRAATSKATTPPKQPKTAPKPKPTSTKPKPMRVRK 179
 QY 1046 PRTTTPPKMTSTMBELNPTSRILAEAMLTQTRPNQTPNSKLVENPKSEDAAGAGETP 1105
 DB 180 PRTTTPPKMTSTMBELNPTSRILAEAMLTQTRPNQTPNSKLVENPKSEDAAGAGETP 239
 QY 1106 HMLLRPHVFMPEVTPDMVLPPEVNOGIIINPMLSDEFNI CNKPEVDGLTTLRNGTLVAF 1165
 DB 240 HMLLRPHVFMPEVTPDMVLPPEVNOGIIINPMLSDEFNI CNKPEVDGLTTLRNGTLVAF 299
 QY 1166 RGHYWMMLSPSPSPSPARITTEVWGISPIDTVFTRCNCEGTFEKKOSQVWRFTNDIKD 1225
 DB 300 RGHYWMMLSPSPSPSPARITTEVWGISPIDTVFTRCNCEGTFEKKOSQVWRFTNDIKD 359
 QY 1226 AGYPKPIFKGFGGLTGOIIVAAALSTAKYKMPESVYFFKRGGSIOQYIYKQEPVQKCPGR 1285
 DB 360 AGYPKPIFKGFGGLTGOIIVAAALSTAKYKMPESVYFFKRGGSIOQYIYKQEPVQKCPGR 419
 QY 1286 PALNTPVYGEWTOVRRRRFERAIGSQTHTRIRIQYSPARLAYODKVLHNEKVSILMRG 1345
 DB 420 PALNTPVYGEWTOVRRRRFERAIGSQTHTRIRIQYSPARLAYODKVLHNEKVSILMRG 479
 QY 1346 LPNVVTSALSLPNTRKPGDYAASKQOYNIIDPSTARAITTRSGQTLKSWYNNCP 1404
 DB 480 LPNVVTSALSLPNTRKPGDYAASKQOYNIIDPSTARAITTRSGQTLKSWYNNCP 538
 RESULT 10
 ABUS3254
 ID ABUS3254 standard; Protein; 513 AA.
 XX AC ABUS3254;
 XX DT 14-APR-2003 (first entry)
 XX DE Human testes-derived DKFZphc03_4019 homologue #3.
 XX KW Human; gene therapy; vaccine; disease treatment; detection.
 XX OS Homo sapiens.
 XX EN WO200112659-A2.
 XX PD 22-FEB-2001.
 XX PF 18-AUG-2000; 2000MO-1B01496.
 XX PR 18-AUG-1999; 99US-0149499.
 XX PR 28-SEP-1999; 99US-0156503.
 XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX PI Wiemann S;
 XX DR WPI; 2001-327840/34.
 XX PT Nucleic acids having the sequences of clones isolated from libraries of

PT different human tissues, useful in recombinant DNA methodologies -
 PS Example III; Page 893; 1095pp; English.
 XX This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention.
 SQ Sequence 513 AA:
 Query Match 36.6%; Score 2757; DB 22; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.4e-140;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 207 RTKKKPTPKPVVDAGSGLDNGDFKTTPTSTTQHNKVSSTKITTAKINRPSLIP 266
 DB 1 RTKKKPTPKPVVDAGSGLDNGDFKTTPTSTTQHNKVSSTKITTAKINRPSLIP 60
 QY 267 NSDTSKETSILTVNKEETVETKETTINKOTSIDGKEKTSKETSKEOSIAKDLAPTSK 326
 DB 61 NSDTSKETSILTVNKEETVETKETTINKOTSIDGKEKTSKETSKEOSIAKDLAPTSK 120
 QY 327 VLAAPTPEAETTTGPAALTPPKBPPTTPPKBPASTTPKEPTTIKSAPTTPKEBAPT 386
 DB 121 VLAAPTPEAETTTGPAALTPPKBPPTTPPKBPASTTPKEPTTIKSAPTTPKEBAPT 180
 QY 387 KSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 446
 DB 181 KSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 240
 QY 447 TPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 506
 DB 241 TPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 300
 QY 507 APTTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 566
 DB 301 APTTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 360
 QY 567 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 626
 DB 361 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 420
 QY 627 TTPKLAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 686
 DB 421 TTPKLAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 480
 QY 687 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 719
 DB 481 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT 513
 RESULT 11
 AAR80041
 ID AAR80041 standard; Protein; 452 AA.
 XX AC AAR80041;
 XX DT 25-MAR-2003 (updated)
 XX DT 10-APR-1996 (first entry)
 XX DE Human megakaryocytopoietin protein.
 XX KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 KW multipotential stem cell.

```

OS Homo sapiens.
XX Key Location/Qualifiers
FH MISC-difference 393..396
FT MISC-difference 444..446 /note="unspecified amino acids"
FT MISC-difference 444..446 /note="unspecified amino acids"
XX MO9523861-A1.
XX
XX PD 08-SEP-1995.
XX
XX PF 06-MAR-1995; 95MO-CN00015.
XX
XX PR 04-MAR-1994; 94CN-0112066.
XX
XX (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
XX
XX Gu X, Han Z, Shen Q;
XX
XX MPI; 1995-320576/41.
XX DR N-PSDB; AAT04546.
XX
XX New haematopoietic cell growth factor - used for treating
XX thrombocytopenia and hematocytopenia
XX
XX Example; Page 23; 36pp; Chinese.
XX
XX This sequence represents the human megakaryocytopoietin (MPO) protein.
XX This sequence was purified using a carrier which can couple wheat germ
XX agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX AAR80039 and AAR80040) were used to produce the amplification primers
XX shown in AAT04544 and AAT04545. The fragments amplified by these primers
XX can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX The MPO cDNA can then be inserted into a plasmid which is used to
XX transform cells to produce MPO. The MPO sequence is capable of promoting
XX colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX and stimulating the proliferation of multipotential stem cells. The
XX factor may be used for treating thrombocytopenia and hematocytopenia.
XX The purification method can be used to isolate MPO from human urine or
XX serum of patients with aplastic anaemia, and from animal blood or urine
XX by radiation exposing the animals to induce aplastic anaemia.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 452 AA;
Query Match 29.2%; Score 2197; DB 16; Length 452;
Best Local Similarity 84.3%; Pred. No. 1.6e-110;
Matches 419; Conservative 3; Mismatches 21; Indels 54; Gaps 2;
QY 1 MAMKTLPIYLLILLSVPIQOVSSODLSGACGCGEGYRDATCNCNDYNCQHYMECCPP 60
DB |||||
QY 61 KKVCTALSLCKGRCFSPERGRCDCDAOCKKYDKCCPDYSPFCAEVHNPTSPSSKAP 120
DB |||||
QY 61 KKVCTALSLCKGRCFSPERGRCDCDAOCKKYDKCCPDYSPFCAEVHNPTSPSSKAP 120
DB |||||
QY 121 PPSGASQTIKSTTKSPKPKPKKKTKVISEBEITEHSHVSENQSSSSSSSSSTIIM 180
DB |||||
QY 121 PPSGASQTIKSTTKSPKPKPKKKTKVISEBEITEHSHVSENQSSSSSSSSSTIIM 180
DB |||||
QY 181 KIKSSKSNANRELQKLVKONKKKRTKKKPPKPPVVDGASGLDNGDPKVTTPDTST 240
DB |||||
QY 157 -----VKONKKKRTKKKPPKPPVVDGASGLDNGDPKVTTPDTST 197
DB |||||
QY 241 TQHNKVTSPKITTAPINPRLSPNSDTSKETSITVNEKETTETKETTNNKQSTNG 300
DB |||||
QY 198 TQHNKVTSPKITTAPINPRLSPNSDTSKETSITVNEKETTETKETTNNKQSTNG 257
DB |||||
QY 301 KKKTTSAKTSQSIKTSKAVDLAFTSKVLAKPAPKAETTTKGALTTPEKPTTPKEPAS 360
DB |||||
QY 258 KKKTTSAKTSQSIKTSKAVDLAFTSKVLAKPAPKAETTTKGALTTPEKPTTPKEPAS 317
DB |||||

```

```

QY 361 TTPKEPTPTTIKSAPTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTKEPAPTTKXP 420
DB |||||
QY 318 TTPKEPTPTTIKSAPTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTKEPAPTTKXP 377
DB |||||
QY 421 APTTTKSAPTTKEPAPTTTPKPPAPTPKEPAPTTTPKEPTPTTTPKEPAPTTKEPATPK 480
DB |||||
QY 378 APTTTKSHPLPRSCXXXCTGP-----TPKEPHPLPRSLHPTTKEPAPTTKE 426
DB |||||
QY 481 EPAPTAKKKPAPTTPE 497
DB |||||
QY 427 EPAPTAKKKPAAPLPLE 443
DB |||||
RESULT 12
AAB60569
ID AAB60569 standard; Protein; 472 AA.
XX
XX AAB60569;
XX
XX 27-APR-2001 (first entry)
XX
XX Bovine MSF orthologue, superficial zone protein (SZP).
XX
XX Bovine; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
XX superficial zone protein; SZP; MSF orthologue; synovial lubricant;
XX osteoarthritis; joint lubrication; osteopathic; antiarthritic.
XX
XX Bos taurus.
XX
XX WO200107068-A1.
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000WO-US20002.
XX
XX 23-JUL-1999; 99US-0145328.
XX
XX 19-JUL-2000; 2000US-0145328.
XX
XX (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Warman ML;
XX
XX MPI; 2001-182721/18.
XX
XX New composition comprising the campodactylly-arthropathy-coxa
XX vara-pericarditis protein in combination with an anesthetic, useful for
XX treating osteoarthritis, or as lubricants of tissue and joints
XX
XX Example 1; Fig 4; 34pp; English.
XX
XX The invention relates to a method of treating osteoarthritis via the
XX administration of a composition comprising the campodactylly-arthropathy-
XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX The composition may further comprise a local anaesthetic. The composition
XX of the invention may be administered via intra-articular or intravenous
XX injection. The human CACP protein is identified in the invention as
XX being megakaryocyte stimulating factor (MSF). The gene encoding
XX CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
XX this gene are responsible for the heritable disorder campodactylly-
XX arthropathy-coxa vara-pericarditis in which patients have synovial
XX hyperplasia without evidence of inflammation. CACP protein (MSF)
XX acts as a synovial lubricant, and can be used to lubricate tissue and
XX joints in the treatment of osteoarthritis. The composition may be
XX applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX loss of range of movement or joint damage). The present sequence
XX represents the bovine orthologue of human MSF, superficial zone
XX protein (SZP).
XX
SQ Sequence 472 AA;
Query Match 22.7%; Score 1707.5; DB 22; Length 472;
Best Local Similarity 79.9%; Pred. No. 3.8e-84;

```

Matches 319; Conservative 24; Mismatches 53; Indels 3; Gaps .3;

QY 1007 PKDRAATSKATTPKQKPTKAPKPKPTSTKCKPTKMPRRVKKPTTPPRK-MTSTMPELNPT 1065
 DB 76 PKGRATNSQVTPPKPKQKPTKAPKPKPTSTKCKPT-PRVKKPTTPPRKTTTSAMBEPPPT 134
 QY 1066 SRIRAMQOTTTPRQGTNSKLVENPKSEDAAGAGEGTTPMLLPHVFMPEVTPDMOYL 1125
 DB 135 S-LPEAMQOTTTPRPTNSSEIIVNSENEDDAAEGKPHMTFPPVLTPIVIGTETII 193
 QY 1126 PRVNOGIIINPMLSDETNINCGRKPEVDGLTLRNGTLVAFRGHYFWMLSPPSPSPARRI 1185
 DB 194 VRGSGQGINPMSEDETNLCGRVVDGLTLRNGIIVAFRGHYFWMMLTPFPPPPRRI 253
 QY 1186 TEWGIPIPIPTVTRFCNCEGKTFFPKDSQVRFNDIKDAGYPPKIRKGGGLTGQIVA 1245
 DB 254 TEWGIPIPIPTVTRFCNCEGKTFFPKDSQVRFNDIKDAGYPLILSKRGGLNGKIVA 313
 QY 1246 ALSTAKYKMPESVYEFKRGGSIOQYIKQEPVOKCPGRPALNYPVYGMTQVRRRRE 1305
 DB 314 ALSLAQYKSRPESVYEFKRGGSVOQYTYKQEPVOKCTGRPALNYSVIGETAOVRRRRE 373
 QY 1306 RAIGPSQHTTIRIOYSPARLAYODKGVLANEVKVSILWRGLPNVVTSAISLPIRKPDGY 1365
 DB 374 RAIGPSQHTTIRIHYTPVRVPYQDKGFLHNEVKVSTLWRGLPNVVTSAISLPIRKPDGY 433
 QY 1366 DYAFSSKQYINIDVPSRTARAITTRSGQTLSSKAYNCP 1404
 DB 434 DYALSKQYINIDVPSRTARAITTRSGQTLSSNTYNCP 472

RESULT 13

AAU11261 standard; Protein; 292 AA.

AAU11261;

12-MAR-2002 (first entry)

Human HAPO polypeptide.

Human; HAPO; stimulating factor; haemopoietic stem ancestral cell; vascular endothelial cell; haemopoietic disease; vascular disease.

Homo sapiens.

CN1312294-A.

12-SEP-2001.

28-FEB-2001; 2001CN-0109083.

28-FEB-2001; 2001CN-0109083.

(HEMA-) INST HEMATOLOGY CHINESE ACAD MEDICAL SCI.

Han Z, Liu Y, Cai Y;

WPI; 2002-018506/03.

N-PSDB; AAS16924.

Human blood and blood vessel cytopoiesis hormone and its preparation -

Claim 1; Page 9 (Disclosure); 20pp; Chinese.

The invention relates to a stimulating factor, HAPO, with the effect of promoting growth of haemopoietic stem ancestral cells and vascular endothelial cells. HAPO is characterised using several pairs of primers specific for HAPO DNA to make a series of gene segments by PCR amplification. An antibody specific to the HAPO polypeptide can be used in the treatment and prevention of haemopoietic and vascular diseases. This sequence represents the HAPO polypeptide of the invention.

SQ Sequence 292 AA;

Query Match 20.5%; Score 1545; DB 23; Length 292;

Best Local Similarity 86.9%; Pred. No. 1.2e-75; Indels 44; Gaps 1;

Matches 292; Conservative 0; Mismatches 0; Indels 44; Gaps 1;

QY 41 DATNCNVCNCOHWECCPDFKRVCTAELSCGRCFESFERGECDCDAQCKYKCCPDY 100
 DB 1 DATNCNVCNCOHWECCPDFKRVCTAELSCGRCFESFERGECDCDAQCKYKCCPDY 60
 QY 101 ESFCAEYHNPSPSSKAPPSGASQTIKSTTRSPKPKKTKYVISEETIEHSV 160
 DB 61 ESFCAEYHNPSPSSKAPPSGASQTIKSTTRSPKPKKTKYVISEETIEHSV 116
 QY 161 SENESSSSSSSSSSSTIKIKSKSAAARELOKLANVKNCKNTKKKTPKPVVD 220
 DB 117 -----KDNKRNKTKKTPKPVVD 136
 QY 221 EAGSLDNGDEKVTTPDSTTOHNVKSTSPKITTAKEINPSPSLPPNSDTSKETSLTVNK 280
 DB 137 EAGSLDNGDEKVTTPDSTTOHNVKSTSPKITTAKEINPSPSLPPNSDTSKETSLTVNK 196
 QY 281 ETVETKETTTNKOJSTDGKEKTSKAKETOSIEKTSKADLAPTSKYLAKEPTPAETTTK 340
 DB 197 ETVETKETTTNKOJSTDGKEKTSKAKETOSIEKTSKADLAPTSKYLAKEPTPAETTTK 256
 QY 341 GPALTTPKEPTTPKPEASTTPKEPTPTTKAPT 376
 DB 257 GPALTTPKEPTTPKPEASTTPKEPTPTTKAPT 292

RESULT 14

AAM24516 standard; Protein; 5179 AA.

AAM24516;

12-OCT-2001 (first entry)

C899P predicted amino acid sequence.

Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic; gene therapy; vaccine; colonic cancer.

Homo sapiens.

WO200149716-A2.

12-JUL-2001.

29-DEC-2000; 2000WO-US35596.

30-DEC-1999; 99US-0476296.

10-JAN-2000; 2000US-0480321.

15-FEB-2000; 2000US-0504629.

06-MAR-2000; 2000US-0519444.

19-MAY-2000; 2000US-0575251.

29-JUN-2000; 2000US-0609448.

28-AUG-2000; 2000US-0649811.

(CORI-) CORIXA CORP.

Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

King GE, Wang T, Jiang Y;

WPI; 2001-441847/47.
 Colon tumor associated proteins and nucleic acids useful for the prevention, diagnosis and treatment of colonic cancer -
 Claim 2; Page 446-462; 472pp; English.
 The present invention describes colon tumour associated proteins (I) and

CC the polynucleotides (II) that encode them. (I) have cytosolic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patient's own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 5179 AA;

Query Match 15 8%; Score 1188; DB 22; Length 5179;

Best Local Similarity 30.9%; Pred. No. 4.7e-55; Indels 279; Gaps 54;

Matches 405; Conservative 87; Mismatches 539;

4 KTLPIYLLLLSVFIQVVSODLSSC--AGRCG-----EGYSRDATCNDYCNQHYMEC 56
 1169 KGRPIY-----EEDLKVCVADKCGCVDETHYPRGASVTEETCSGC-V 1213
 57 CPDFKRCVTAELSCGRCFSEFBERG---RECCDCAQCKYDKCC---PDYSFCME 106
 1214 TNSGQVCRPE---EGKINOTODGAFCEYICGPRGVEKHPNICSITRSTLTFTT 1270
 107 VNPPTSPS-SKKAPPSGASOTIKSTKSPKPKKTKKVISEETEEH---SVSN 163
 1271 IYLPPTTPTSTTTTTTPTSTVST---PK-----LCCLMSDWNEHPSGSD 1320
 164 QSSSSSSSSSSSTIWKIKSSKNANRE-----LOKLLKYDNKNTKKKPPKPV 218
 1321 GDRPPDGVCGAPBDI-ECRSYKDPHLSLEHGQKQCVSVGFIKME----- 1368
 219 VDEAGSGLDN--GDPKVTTPDTSTQHNKYSTSPK-ITTKAPINRPSLPPNSDTSKETS 275
 1369 -DQFGNGPFGLCYDYKIRV-----NCCWPMDCITTPSPPTTTPSPPTTTTLPPT 1419
 276 LTVNKETVETKETTNNKQSTDGKETSATAKOSIEKTSAKDLAPTSKVLAKPTPKA 335
 1420 TTPSPPTT-----TTTTPTTTPSPPTTTTTP-----LPTT-----TPSPPI 1458
 336 ETTTGAPALTTPKEPT-----PTPKAPASTTPKEPTTTIKAP--TPKEP-----A 382
 1455 GTTTTTPPTTTPSPPTTTPSPPTTTPPTTTTTPPTTTPSPPTTTPPTTTPPTT 1518
 383 PTTTKAP--TPPKAPATTKEPAPTTPKEP-----APTTKBAPTTPKAP--TTP 432
 1519 PTTTSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1578
 433 KEPAATPKKAPATTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 485
 1579 SPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1638
 486 APKKAPATTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPATTTPKAPATTTPKAP 545
 1639 TTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPT 1698
 546 P---TPPKSPPTTTPKEPAPTTPKEPAPTTPKAPATTTPKEPAPTTPKAPATTTPKAP 602
 1699 PPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1757

QY 603 TAPKEPAPTTP-----KEPAPTTP----- 621
 DB 1758 TTPLEPSTTPPTSPPTTTPPTTTPCVPLCNWTGMGLDSKPNFHKRGDTELIGDVGRCW 1817
 QY 622 -----KKLP----- 626
 DB 1818 AANISCRATMPDVPDVGIGLGQVVDVSGVLCCKNEQDKPGVIMAPACLANEINVQCE 1877
 QY 627 --TPKEAPATTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 683
 DB 1878 CVTQPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1937
 QY 684 TPKEP-----APTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 737
 DB 1938 GTQPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1997
 QY 738 TKEPTSTSDKRAP--TPKGTAPATTTPKEPAPTTPKEPAPTTPKGTAPATTTPKEPAPTTP 796
 DB 1998 TTPPTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2056
 QY 797 PAPKEAPATTTPKGTSTSDKRAP--TPKEPAPTTPKEPAPTTPKAPATTTPPTPTPT 855
 DB 2057 PTTTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2115
 QY 856 EVSTPTTTPKEPTTHKSPDSSTPELSAETPKALENSKEP-----CVPT--TKTP 904
 DB 2116 PTPPTTPGTQPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2172
 QY 905 AATKEPMTTAPKDKTERDLR---TTP--ETTAPAKMT---KEPATTEKTESKITAT 956
 DB 2173 TTTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2232
 QY 957 TTVQVSTTTPQTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2287
 DB 2288 TTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2347
 QY 1072 MLOQTTT--PNOTPNSKLVENPKSEDAGAGETPHMLLRPHVEMEVTP 1120
 DB 2348 PTPPTQPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 2392

RESULT 15
 ABP55365
 ID ABP55365 standard; Protein, 5179 AA.
 XX
 AC ABP55365;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Human colon tumour protein for clone C899P SEQ ID NO.1068.
 KM Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
 KW tumour; immune response; immunostimulant; cytostatic; vaccine.
 OS Homo sapiens.
 PN MO200283070-A2.
 PD 24-OCT-2002.
 PF 09-APR-2002; 2002MO-US11475.
 PR 10-APR-2001; 2001US-0833263.
 PR 03-AUG-2001; 2001US-0922217.
 PR 19-DEC-2001; 2001US-0025380.
 XX
 XX (CORI-) CORIXA CORP.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:31:27 : Search time 53 Seconds
(without alignments)
6835.960 Million cell updates/sec

Title: US-09-556-246-1
Perfect score: 7526
Sequence: 1 MAWKLPITVILLSLVFVILQ.....ARATRRSGQTLSKWNVNC 1404

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL.23.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_virus.*
 - 16: sp_bacteriopl.*
 - 17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	4	Q92954
2	7457	99.6	1404	4	Q9BX49
3	3968	53.0	1054	11	Q9JW99
4	1713.5	22.8	401	6	077765
5	1215	13.1	1152	5	Q9VR49
6	982.5	12.9	3458	5	Q8IR52
7	974	12.9	3458	5	Q8IR52
8	968.5	12.0	3432	5	Q8IR51
9	902	12.0	1795	5	Q76894
10	896.5	11.9	1349	4	Q8MWQ4
11	863.5	11.5	1489	10	Q96449
12	846	11.2	1274	5	Q20007
13	843.5	11.2	934	5	Q9VVG2
14	829	11.0	1607	10	Q8H6Q5
15	816	10.8	1315	10	Q9SPM0
16	812	10.8	2187	11	P70670

17	811.5	10.8	1480	10	Q9LIE8	Q9LIE8 arabidopsis
18	784	10.4	1188	10	Q41805	Q41805 zea mays (m
19	782.5	10.4	2112	5	Q9VLE9	Q9VLE9 drosophila
20	767.5	10.2	1009	10	Q8L685	Q8L685 volvox cart
21	746	9.9	1254	5	Q94185	Q94185 caenorhabdi
22	715	9.5	763	2	Q9XDH2	Q9XDH2 mycobacteri
23	710	9.4	4969	11	Q8CF91	Q8CF91 mus musculu
24	710	9.4	5165	11	Q8CF92	Q8CF92 mus musculu
25	709.5	9.4	18519	5	Q8ISF6	Q8ISF6 caenorhabdi
26	709.5	9.4	18514	5	Q8ISF7	Q8ISF7 caenorhabdi
27	703	9.3	10578	5	Q8ISF5	Q8ISF5 caenorhabdi
28	695	9.2	1151	13	Q57580	Q57580 gallus gall
29	678	9.0	2090	4	Q96QC2	Q96QC2 homo sapien
30	677	9.0	598	16	Q8VKN7	Q8VKN7 mycobacteri
31	677	9.0	2114	5	Q8L018	Q8L018 drosophila
32	675.5	9.0	2768	5	Q9VC00	Q9VC00 drosophila
33	674.5	9.0	2089	4	Q14676	Q14676 homo sapien
34	669.5	8.9	2284	5	Q9VPG1	Q9VPG1 drosophila
35	666.5	8.9	7962	4	Q10465	Q10465 homo sapien
36	665.5	8.8	34350	4	Q8WZ42	Q8WZ42 homo sapien
37	643.5	8.6	990	13	Q91803	Q91803 xenopus lae
38	636	8.5	6632	5	Q17362	Q17362 caenorhabdi
39	635	8.4	1720	5	Q81486	Q81486 plasmodium
40	633	8.4	489	10	Q41707	Q41707 vigna ungu
41	632	8.4	761	10	Q92010	Q92010 arabidopsis
42	631	8.4	2344	5	Q9N3Y8	Q9N3Y8 caenorhabdi
43	629	8.4	6632	5	Q01761	Q01761 caenorhabdi
44	627.5	8.3	971	5	Q9XV54	Q9XV54 caenorhabdi
45	622.5	8.3	839	16	Q9RX57	Q9RX57 deinococcus

ALIGNMENTS

RESULT 1
ID Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Megakaryocyte stimulating factor.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhattacharya S., Kriz R., Hewick R., Clark S.C.;
RT "Purification, Biochemical Characterization, and Cloning of a Novel
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
RT Stimulating Activity.";
RL Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Werberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scaltreto J., Kelleher K., Preisner K., Kriz R.,
RA Jacobs K., Turner K.;
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";
RL (In) Preisner K.T., Rosenblatt S., Kost C., Wegernoff J.,
RL Mosher D.F. (eds.);
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52,
RL Elsevier Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhattacharya S., Kriz R., Hewick R., Clark S.C.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.


```
Db 61 KRVCSPBELSCGRGCFESFARGBRECDQCKQYGCADYDSFCEVHNSTP-PSKTAAP 119
Qy 121 PPSGASQITKSTKTRSPKPKKTKKATYIESEETIEHSVSENQSSSSSSSSSTW 180
Db 120 TPAGASDITKSTKTRSPKSPKT-TRITKYIESEBELTIEHSDBENQ--SSSSSSSSSTIR 176
Qy 181 KIKSSKNSANRELOKKLVKONKKNRTKKKPTPKPVVDEAGSLDNGDFKVT--TPDT 238
Db 177 KIKSSKNS-ANRELOKKNVNDKONKKNRTKKKPNPEPPVNDERGSLDNGERKLTTPPPDP 235
Qy 239 STTOHKNVSTSPKITTAKPINRPSLPNSDSKESLTVNKETVETKETTNNKQST 298
Db 236 PTHSHSKVATSPKTTAAKPVTPKPSLAPNSETSKASLANKETTVETKETTATNKQSSA 295
Qy 299 DGEKTSKAKTOSIEKTSADOLAFSTKVLAKPPTKAEITTKGPAITPKPEPTTPKEP 358
Db 296 -SKKKTTSVKETRSKETSDDV-----EP 319
Qy 359 ASTPKPEPTPTIKSAPTTPKEBAPTTSKAPTTPKEBAPTTPKEBAPTTPK 418
Db 320 TSTTPK-----NSAPTTPKKPV-TTKESKFLD-----LPQEPPTAK 357
Qy 419 EPAPTTSKAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 478
Db 358 EPPPTTKKPEPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 417
Qy 479 PKBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPK 538
Db 418 PKBAPTTPKPEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 461
Qy 539 PTTKSAAPTTPKEBAPTTPKEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 598
Db 462 -----EPTTKPEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 512
Qy 599 KPAAPTAPKBAAPTTPKETAAPTTPKKLTPTTPEKLAAPTTPKEBAPTTPPEBAPT 658
Db 513 ----- 512
Qy 659 TTBEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPK 718
Db 513 -----TPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 539
Qy 719 EPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 778
Db 540 EPEBAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 570
Qy 779 KGIAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 838
Db 571 -----KEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 609
Qy 839 PKBAPTTPPEPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 898
Db 610 PKBAPTTP----- 617
Qy 899 PTTKPAATKEMTTAKDKTTERDLRTPEPTTAAKMTKETATTTETTESKITATTT 958
Db 618 -----TSPKT----- 622
Qy 959 QVSTTTODTTPFKITLTKTTTLAPKVTTKKITTTEINMKPEBETAKPKDRAATNSKAT- 1017
Db 623 -----TLKATTLAPKVTAPAE-----EIQKPEBETTPASSEDSDSKTTL 662
Qy 1018 -----TPKP-QKTKAPKPKPTSTKPKKTPMRYVAKKPTTPPKRMSTMBELNPTSIA 1069
Db 663 KPOKPTKAPKPKPTSTKPKKPTSTKPKKPT-PTKTRKPTTPAPLKTTSATBELNPTT-P 719
Qy 1070 EAMQTTTRPMQTPSKLVEVNPKSESDAGABGEMHMLRPHVMEVPTDMOVLPRVP 1129
Db 720 EVMPLPTTTPKQTPPEPTAPVNDHEDADGGEKXP-LIPGPVLPFPAIDETDLAAGRL 778
Qy 1130 NOGIINMLSDETNINCAGVNDGLTTLRNGTLVFRGHYFVMLSPFSPSPARRITEVW 1189
Db 1130 NOGIINMLSDETNINCAGVNDGLTTLRNGTLVFRGHYFVMLSPFSPSPARRITEVW 1189
```

```
Db 779 NRGININPMPEDETNLNGKPEVDGLTTLRNGTLVFRGHYFVMLNPRPSPARRITEVW 838
Qy 1190 GISPIDITVFRKNCCEKTEPFKDSOYMRFTNDIKDGYRPIKFGGLTGOVVALST 1249
Db 839 GISPIDITVFRKNCCEKTEPFKDSOYMRFTNDVDVDPYKQIVKGGLTGKIVAAUST 898
Qy 1250 AKYKMWPESSVYFFRGSGSIQYIYKQEPVOKCGRBPALANYPVYGMTQVARRRFEPAIG 1309
Db 899 AKYKDRPESVYFFRGSGSIQYIYKQEPVOKCGRBPALANYPVYGMTQVARRRFEPAIG 958
Qy 1310 PSQHTTRIOYS-PARLAYODKGVLANEVKYSILMRGLPNVTIJAISLPINRKPDGDY 1368
Db 959 PFQHTTRIHYSVMRVSYODKGLHNEVKYSTWMRGFPNVVTSAILPLNIRKPDGDY 1018
Qy 1369 AFSKQYVNIIDVPSRTARITTRSGOTLSKVWVNC 1404
Db 1019 AFSKQYVNIIDVPSRTARITTRSGOTLSKVWVNC 1054

RESULT 4
ID 07765 PRELIMINARY; PRT; 401 AA.
AC 07765;
DT 01-NOV-1998 (TRENBLREL. 08, Created)
DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Superficial zone protein (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
CX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99171663; PubMed=920774;
RA Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B.,
RA Kuettnner K.E., Caterson B.;
RT "Articular cartilage superficial zone protein (SZP) is homologous to
RT megakaryocyte stimulating factor precursor and is a multifunctional
RT proteoglycan with potential growth-promoting, cytoprotective, and
RT lubricating properties in cartilage metabolism."
RL Biochem. Biophys. Res. Commun. 254:535-541 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99171663; PubMed=10073655;
RA Schumacher B.L., Hughes C.E., Kuettnner K.E., Caterson B.,
RA Aydelotte M.B.;
RT "Immunodetection and partial cDNA sequence of the proteoglycan,
RT superficial zone protein, synthesized by cells lining synovial
RT joints."
RL J. Orthop. Res. 17:110-120 (1999).
DR EMBL; AF056218; AAD13404.1; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF00045; hemopexin; 2.
DR SMART; SM01217; PRICHEXTENS.
DR SMART; SM00120; HX; 2.
DR PROSITE; PS00024; HEMOPEXIN; 1.
FT NON TER 1
SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFB73D7 CRC64;

Query Match 22.8%; Score 1713.5; DB 6; Length 401;
Best Local Similarity 80.2%; Pred. No. 5e-86;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;

Qy 1007 PKDASTSKATTPKOKPTAKPKKPTSTKPKKTPMRYVAKKPTTPPKRMSTMBELNPT 1065
Db 5 PKGSAITSQVTPKQKPTAKPKKPTSTKPKKPT-PRVRKPTTPPTTTSAMBEPTPT 63
Qy 1066 SRIEAMQTTTRNQTPNSKLVEVNPKSESDAGABGEMHMLRPHVMEVPTDMOVL 1125
Db 1066 SRIEAMQTTTRNQTPNSKLVEVNPKSESDAGABGEMHMLRPHVMEVPTDMOVL 1125
```

DB 64 S-LPEAMLQTTTPTPTPNSSEIDVNSSENGDAEAGEKPHMIFRPVLTPIVPGTEII 122
QY 1126 PRVPGGIIINPMLSETINCNKGPDLGTLTNGTLVARGHYFMMLSFSPSPARRI 1185
DB 123 VGPSPGCGFNGPMSDBTLCNCRPVDGLTTLNGLTVARGHYFMMLTPTPTPPRR 182
QY 1186 TEWVGJPSPIIDVFTCNCEGKTEFPFADSOYRFTNDIKDAGYKPKIPKFGGLTQOIVA 1245
DB 183 TEWVGJPSPIIDVFTCNCEGKTEFPFADSOYRFTNDIKDAGYKPKIPKFGGLTQOIVA 242
QY 1246 ALSTAKYKMPSESVYFPKGGSG10QY1YKQEPVQKCPGRPALNYPVGEVMTQVRRRRE 1305
DB 243 ALSIAGYKSRPESVYFPKGGSGVQOYTYKQEPQCKTGRRPALNYSVGEVTAQVRRRRE 302
QY 1306 RAIGPQGTIRIRQYSPARALYDQKVLNENKVSILMGLPNVVTSAISLPITRPPDGY 1365
DB 303 RAIGPQGTIRIRHYTPVPRVYDQKGLNENKVSILMGLPNVVTSAISLPITRPPDGY 362
QY 1366 DYAFSPKDOYVNDVPSRTARATTTSSGOTLSKVMNCP 1404
DB 363 DYAFSPKDOYVNDVPSRTARATTTSSGOTLSKVMNCP 401
RESULT 5
QYVR49 PRELIMINARY; PRT: 1152 AA.
AC QYVR49
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE CG3047 protein.
OS SGSI OR CG3047.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutcliffe R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA George R.A., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriell J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Banu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kravitz C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mleishina N.V., Moberg C., Morris J., Moshiref A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pitsman K.A., Nixon K., Nusseken D.R., Pacleeb J.M.,
RA Palazolo M., Pitsman K.A., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier L., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskae R., Tector C., Turner R., Venter G.M., Wang A.H., Wang X.,
Wang Z.-Y., Weissman D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsner V., Doup L.E., Doyle C., Drenek D., Farfan D.,
RA Ferriera S., Friese E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshiref A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleeb J., Paragae V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pitsman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mistr S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Flybase;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB003575; AA05957.2; -
DR Flybase; FB00003372; Sg91.
DR Interpro; IPR002965; P_rich_extensan.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1152 AA; 118473 MW; 7538C0E972571F06 CRC64;
Query Match 16.1%; Score 1215; DB 5; Length 1152;
Best Local Similarity 30.4%; Freq. No. 2,4e-58;
Matches 337; Conservative 128; Mismatches 493; Indels 150; Gaps 34;
QY 98 PYSEFCAEVNHPPTPPSKKAPPPSGASQRTKS-----TTKSPPPNKKKTX 146
DB 21 PPVPGCDTDSPTTTKPRQKTKRKRTRTKTKRPPRKTMTTKATKRTTKTR 60
QY 147 K--VIESEITEHSVSENOSSSS--SSSSSSSTTWIKIKSSKNSAANRELQKKLYK 201
DB 81 RRPPTKPTDPTDTSITTTGACCTCSDDRTASTDST-----TDRTYVNTDWTPLCTD 135
QY 202 DNKKRRTKKKP--KKPPVVDASGDLNGDKPKYTPDSTTQHNKVSPPKITTAKPN 259
DB 136 TPPCTCSSESSSTAIDSSPCIDTSYVIPSPCTQETTPPTPSTGCTGTPTCTCAQT 195
QY 260 PPSPSPNSDTSKETSLSLVNKEETVER--KETTITKQSTDEKTKTSKAKTOSIEK--T 316
DB 196 PRST--TTTSTSRPTTTTTPRSTTTTTSRPTTTTSTTT-----TTTRRPTTTT 248
QY 317 SAKDLAPISKVLAKPTPAETTTKGPALTTPEKPTTTPKEPASTTTPKEPPTTKSAPT 376
DB 249 TTSTCAPTT--TTPRSTTTTTSRPTTTTTPCTTTTSCSPTRTPPS--TTTSTSRPT 304
QY 377 TKKEPAPTTTS-----APTTKEPAPTTTPKEPAPTTPKEPAPTTPPTTKSAP 429

Db 305 T-TTPRCCTTSTCTRRPPTTTPSTTTCTSGPTTTTPTSTTTCTSGPTTTTPTSTT 363
Qy 430 TTPPK-PAPTTPK-----PAPTPKAPATTTPKAPTTPPKAPPT-TPKAPATT 479
Db 364 TTTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSGPTTTT 423
Qy 480 KEPAPATKAPATTTPK-----PAPTPKAPATTTPKAPTTPKAPTTPKAPTTPKAPT 531
Db 424 RSTTTTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSGPT 483
Qy 532 TTPKAPATTTPKAP-TPKAPPTTPKAPATTTPK-----PAPTTPKAPATT 581
Db 484 TTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTST 543
Qy 582 KEPAPTTKAPATTTPKAPAPAK-----PAPTPKAPATTTPKAPTTPKAPTTPKAPT 633
Db 544 SGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRST 603
Qy 634 PTPPKAPATTTPKAPATTTPKAPPTTPK-----PAPTPKAPAPNTPKAPATT 685
Db 604 TTTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTP 663
Qy 686 KEPAPTTKAPATTTPKAPT-----APTTPKGIATTPKAPATTTPKAPTTPKAPT 737
Db 664 RSTTTTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPT 723
Qy 738 TKEPTSTSDK---PAPTPKGIATTPKAPATTTPK-----PAPTPKGIATTP 786
Db 724 TTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTST 783
Qy 787 KEPAPTTPK-----PAPELAPTTTKGPTST-----SDKAPATTTPKETA 827
Db 784 SGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRST 843
Qy 828 PTPPKAPATTTPKAPATTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 887
Db 844 TTTSGPTTTTTPRSTTTTCTSGPTTTTTPRSTTTTCTSGPTTT-TPRSTTTTCTSGPTTT 901
Qy 888 ALENSPK-EPGVPTTKPAATKPEMTTAKOKTERDRTTETTTAAPTTPKAPTTPKAPTTPKAPT 946
Db 902 TTPRSTTTTCTSGPTTTTTPRSTTTTCTSGPTTTTTPRSTTTTCTSGPTTTTTPRSTTTTTP 957
Qy 947 KTSKSKITATTTQVSTTTODTTPPKITLTKTTTLAPVTTTUKITTELMNKPEETA- 1005
Db 958 STSPATTTTPRSTTTTSTSGPTTTTTPRSTTTTSTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTT 1015
Qy 1006 KPKORATNSKATTPKAPK-PTKAPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT 1064
Db 1016 RSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTP 1075
Qy 1065 TSRIAEAMLQTTT-----PQOT 1083
Db 1076 S---ASPTTTPPTTTPRCPCHPOPPVQIP 1100

RESULT 6
Q9N4S7 PRELIMINARY; PRT, 1079 AA.
AC Q9N4S7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Y51B11A.1 protein.
GN Y51B11A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;

RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Cotton M.;
RT "The sequence of C. elegans cosmid Y51B11A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006797; AAFC0743.1;
DR Wormpep; Y51B11A.1; CE22326.
DR Interpro; IPR002965; P. rich. extensn.
DR PRINTS; PR01217; PRICHEXTNSN.
SQ SEQUENCE 1079 AA; 110532 MW; 8BDE3824CF80CA1 CRC64;

Query Match. 13.1%; Score 982.5; DB:5; Length 1079;
Best Local Similarity 29.6%; Pred. No. 1.1e-45;
Matches 332; Conservative 131; Mismatches 481; Indels 177; Gaps 44;

Qy 195 QKLVKNDKNNKRRKTKKPPVVDAGSLDNGDFV-----TTPDTSTQ-HNKV 246
Db 27 QKELEIDCRANQPPHMLP-----STLISVDNETBSTVLASTPSSSTPIKETT 77
Qy 247 STSPKITTAKENDRPSLPNSDTSKETSIVNKETT-VETKETTNNKQSTDGKETT 305
Db 78 TTAETTSTP-----PSSSTTPVQTTTAAETTSTAPSSSTTPVQTT-----TT 124
Qy 306 SAKTOSIEKTSADLAPTSIVLAKTPKATT-TPK-ALTTPKEPTTPKAPTTP 363
Db 125 TAPETSTPEPSSS-----TSPVQTTTAAETSTEAAPSSSTTPVQTTTAAETTSTEP 180
Qy 364 KEPPTTIKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPT 423
Db 181 PSSSTPVQTTTAAETTSTPEPSSSTTPVQTTTAAPE---TTSSTEP-PSSSTTPVQTT 236
Qy 424 TTKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPT 483
Db 237 TTTTAAETSTPEPSSSTTPVQTTTAAETTSTEP-PSSSTTPVQTTTAAETTSTEP 295
Qy 484 PTAKKAPATTTPKAPATTTPK-----APTTPKPSPTPKAPATTTPKS 529
Db 296 PSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTT 355
Qy 530 APTTTK--EPAPTTTKAPATTTPKESPTTPK--APTTPKAPATTTPKAPATTTPKAPAPT 586
Db 356 APETTRTEPSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTT---APETSTSTEP 411
Qy 587 TTPKAPATTTPKAPAPAKEP--APTTPKETAPTPKLPTTPKLAAPTTPKAPT 644
Db 412 SSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTTAAETTSTEP---PSSSTTPVQTT 467
Qy 645 PEELAPTTPEPTTPTEBEAPATTPKAAAPTTPK--APTTPKAPATTTPKAPATTTPK 702
Db 468 TTPAETTSTEP-PSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTTAAPE-TTSTES 525
Qy 703 ETATTPKGIATTPKAPATTTPKAPKAPKELAPTTTEPTSTSDK--APTTPKGIATTP 761
Db 526 PSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTT 585
Qy 762 PKKAPATTTPKAP--APTTPKGIATTPKAPATTTPKAPKAPKELAPTTTKGPTSTSDK- 818
Db 586 APE---TTSSTEPSSSTTPVQTTTAAETTSTPEPSSSTTPVQTTTAAETTSTEPSS 642
Qy 819 APTTPKETAAPTTPKAPATTTPK---PAPTTPETTPPTS-----EVSPTTPKE 865
Db 643 SSTTPVQTTTAAETTSTPEPSSSTTPVQTTTAAETTSTPEPSSSNTTPVQTTTAAPE 702

QY 866 PTHHSKPESTBELSAEPT--PKALENSKEPGEVPTTKTPAATKPEMTTAKDTEED 923
 DB 703 ETTSTPSSSTPVQTTTATETSTPSSSTTPVQTTTATETSTPSSSTTP 762
 QY 924 LRTT---PETHAARMTKEATTEKTESKITAATTOVISTTODTTPKITTLLKT 979
 DB 763 VQTTTATETSTPSSSTTPVQTTTATETSTPSSSTTPVQTTTATETSTPSSSTTP 820
 QY 980 TLAPKTTTKITTEINNKPEETA-KPKDRAATNSKATPKKQKTKAPKKTSTK 1038
 DB 821 STPSSSTTPVQTTTATETSTPSSSTTPVQTTTATETSTPSSSTTP 878
 QY 1039 TWPVAKPTTPTPRKATSTMPELNPTSAEAMLOTTTPRNPOTPSKLVENPKSEDAG 1098
 DB 879 QTTTATETSTPSSSTTPVQTTTATETSTPSSSTTPVQTTTATETSTPSSSTTP 932
 QY 1099 GABG-ETPMMLRPHVPMPEVTP--DMOYLPV-----PNOGIITNMLSDFTN 1144
 DB 933 ETTSTPSSSTTPVQTTTATETSTPSSSTTPVQTTTATETSTPSSSTTP 992
 QY 1145 -----ICNGKPVQDLTLNGLTVAFRGHVPMMS-----PSPSS 1180
 DB 993 ANSFVSTPFGICTTALTCTYSEGGISNL-NATL-----FTGLSDGSSIDLPE---- 1040
 QY 1181 PARITVWGPISPIDTFTFRNCSEKTFEFDKSOYWRFTN 1221
 DB 1041 -----YVPTGLE-IMPEINCEGK-----WSYVN 1064

RESULT 7

ID 08IR52 PRELIMINARY; PRT: 3458 AA.
 AC 08IR52;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE CG32602-PA.
 GN CG32602.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Sutcliffe R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abryl J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benoe P.V., Bertram B.P., Bhandari D., Bolhakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher S., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Garg J., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mikhlin N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saundere R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B.C., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svitek R., Tector R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zhou L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreenek D., Farfan D.,
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hosteln D., Howland T.J.,
 RA Idegawa C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragav V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Mathews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celniker S.E.,
 RA Clamp W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RA Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB003495; J000000000.1;
 SQ SEQUENCE 3458 AA; 352129 MW; D532F3D9561B109E CRC64;

Query Match 12.9%; Score 974; DB 5; Length 3458;
 Best Local Similarity 33.0%; Pred. No. 16-44;
 Matches 409; Conservative 69; Mismatches 488; Indels 272; Gaps 65;
 QY 110 PPSPPSSKKA--PPSGAS--QTIKSTTKRPPKPKPKKKTKV/IESEITEHSVSENO 164
 DB 2263 PTEGSAKPTTKAPBGTAKPTTKAPBGTSAKPTTKLP-----EGTAKPTTKLP 2316
 QY 165 ESSSSSSSSSSSSSTTWIKSSRNSANRELQKLVKYNKRNKRTKKKPPKPPVDEAGS 224
 DB 2317 EGTSAKPTT-----LKPTKGTTA-----KPTTLKPTBGRRTAKPTTKLP----- 2356
 QY 225 GIDNGDFKTTTDTST--TTQHNKYSTSPKITTAP-----INRPSLPPNSDTSKETSL 276
 DB 2357 -----TGTSAKPTTKLPTEGTSKAPPTTKLPTEGTTAKPTTKLPBGTAKPTTL 2406
 QY 277 -----TVNKETTV--TKETT-----TNKQF--STDGKKT-----TSAKET--OSIEKTS 318
 DB 2407 KPTBGTSAKPTTKLPKGTAKPTTKLPBGTSAKPTTKLPBGTAKPTTKLPBGTSA 2466
 QY 319 K--DLAPTSKVLAKPT--PKAETTKGPALTPPKPPPTTPKEPASTTPKEPT---PTT 370
 DB 2467 KPTTLKPTBGTAKPTTKLPBGTAKPTTKLPBGTAKPTTKLPBGTAKPTTKLPBGTSA 2522

QY 371 IKSAPTTKEBA---PTTKSAPTTKEBAPTTTKEBA---PTTKBAPTTTKEBAPT 424
 Db 2523 LKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGT 2582
 QY 425 TKSAPTTKEBAPTTTPK---KPAPTTKEBAPTTTPKPTT---PTTKBAPTTTKEBAPT 477
 Db 2583 AK--PTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTL 2640
 QY 478 TPKEBAPPAKPAPTTKEBAPTTTKEBAPTTTKEBS---PTTKBAPTTTSAAPT 534
 Db 2641 KPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGT 2699
 QY 535 KEBA---PTTKSAPTTKEBAPTTTKEBA---PTTKBAPTTTPK---KP 576
 Db 2700 AKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTAKPTTLK 2759
 QY 577 A-----PTTKBAPTTTPK---EPAPTTTKEBA---PTAKBAPTTTKEBAPT---P 621
 Db 2760 TKGTSAPTTTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKP 2819
 QY 622 KKLPTT---TPKLAFTTPEKBAPTT-----PEBLAFT---TPBPTTTPPEP 664
 Db 2820 TTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTB 2879
 QY 665 APPTPKAAAPTTKEBAPTTTKEBAPTTTKEBAPTTTKEBAPT---P 714
 Db 2880 TTKAPTTTLKPTBGTSAKPTTLKPTBGTAK---PTTLKPTBGTAKPTTLKPTBGTSAK 2936
 QY 715 TTLKBAPTTPK---KEBAPKELAPTTTKEPTSTSDKPA-----PTTKPGTAP 759
 Db 2937 TTLKPTBGTAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTB 2996
 QY 760 TTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAPTTK---KPAKELAPTTTKEPTSTTS 815
 Db 2397 TTKA---PTTLK---PTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGT 3050
 QY 816 DKPA-----PTTKBAPTTTKEBAPTT---PKKAPPTTPEPTTSTSTST 862
 Db 3051 AKPTTLKPTBGTSAKPTTLKPTBGTAK---PTTLKPTBGTSAKPTTLKPTBGTAKPT 3107
 QY 863 TK-----EPPTTKHSPESTPELSAETP---KALENSPKBPGV-----PTTKPATAP 909
 Db 3108 LKPTBGTSAKPTTLKPTBGT---KPTBGT---TAKPTTLKPTBGTAKPTTLKPTBGTAKPT 3162
 QY 910 EMTTAAKPTTERTDRTTPTTT-----AAPTKEKATTTTKEPTSTKIT---ATTTOV 960
 Db 3163 TEGTSAPKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSA 3222
 QY 961 TSTTODTTPKPTTL-----KTTTLAPKVTTTK---TTTTTEIMNKKEBTAAPKDR 1010
 Db 3223 TTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTSAKPTTLKPTD 3281
 QY 1011 ATNSKATTPKPKOKPTKAPK---KPT---STKKPKTMP---RVKPKPTTTPRKMSTWMP 1060
 Db 3282 GTTAKPTTLKPTBGTSAKPTTLKPTBGTAKPTTLKPTBGTAKP---TTLKPTBGTSAKPT 3340
 QY 1061 ELNPTSRALAMLOTTTRPNOTPNKSVAVNPKSBDGAGAGETPPHMLRPHVPMPEVTP 1120
 Db 3341 TLKPTT---GTTAKPTTLKPTBGTSAKPTTLK---TEGTT---AKPTTLKPTBGT 3387
 QY 1121 DMDYLRVNVNPGIINPMLSDPTNINCNKCPVDGLTTLR 1158
 Db 3388 SAKPTTLKPTBGTAKP---TTL---KPTDGTTPPR 3417

RESULT 8

Q8IR51
 ID Q8IR51 PRELIMINARY; PRT: 3432 AA.
 AC Q8IR51;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE CG32602-PB.
 GN CG32602.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyzeta;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hinkley R.A., Gale R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Abell J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Metkulov G., Milshina N.V., Nobayri G., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Neilson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rehner K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimlos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgerger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jaitai M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phuanavong S., Peltman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.B., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnitsk F., Whitfield E.,

[illegible]

Qy	819	A-----PTTPEKLPPTPKCEBAPT--PKCBAPTPEPTPTSEVSPPTTK-	864
Db	3028	TTLKPTEGTSAKPTTLKPTEGTAK---PTTLKPTEGTSAKPTTLKPTEGTAKPTTLKP	3084
Qy	865	-----EPTTIHKSPEDESTBELSAEPTP-KALENSPREPV-----PTTKPATKPEMT	912
Db	3085	TEGTSAPPTLK--KPTEGT---TAKPTTLKPTEGTAKPTTLKPTEGTAKPTTLKPTEG	3139
Qy	913	TTADKXTTEBDRPTPTTT-----AAPKTKATTTKTEKTEKII-ATTQVTS	963
Db	3140	TSAPPTTLKPTEGTAKPTTLKPTEGTSAKPTTLKPTEGTAKPTTLKPTEGTSAOPTL	3199
Qy	964	TTQDPTPEFKITTL-----KTTTALPKYTTTK--TITTEIMNKPEETAKPKDATN	1013
Db	3200	KPTDGTAKPTTLKPTEGTSAKPTTLKPTEGTAKPTTLKPTEGTSAOPTTLKPTD-GTT	3258
Qy	1014	SKATTPPKPKTKAPK---KPT--STKKPKTM-----RVAKPKTTPPTPTMTSTMP-ELN	1063
Db	3259	AKPTTLKPTEGTSAKPTTLKPTEGTAKPTTLKPTEGTAKP--TTLKPTEGTSAKPTTLK	3317
Qy	1064	PTSLIAAMQTTTRPNOTNSKLVENPNSBDAGAEGETPHMLRPHVMEVTPDMD	1123
Db	3318	PTE--GTTAKPTTLKPTEGTSAKPTTLKP-----TEGTT-----AKPTTLKPTEGTSAK	3364
Qy	1124	YLPRVPNOGIIINPMLSEDTNIGCKGKVPDGLTTLR	1158
Db	3365	PTTLKPTEGTAKP-----TTL---KPTDGTTEBR	3391
RESULT 9			
ID	076894	PRELIMINARY;	PRT; 1795 AA.
AC	076894		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	EG:5667.1 protein.		
GN	EG:5667.1 OR CG14796.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RA	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RC	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos L.G.,		
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Plamnoch C., Baldwin D.,		
RA	Ballew R.M., Baasu A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotlier P.,		
RA	Burtis K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablo E., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,		
RA	Jatail M., Kalush F., Karpen G.H., Ke Z., Kemnitz J.A., Ketchum K.A.,		
RA	Klammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		

R	A	Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Pacle J.M.,
R	A	Palazzo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,
R	A	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
R	A	Shue B.C., Sliden-Klamos I., Simpson M., Skupski M.P., Smith T.,
R	A	Spiet E., Spieding A.C., Stapleton M., Strong R., Sun E.,
R	A	Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
R	A	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
R	A	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
R	A	Ye J., Yeh R.-F., Zaeri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
R	A	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
R	A	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
R	T	"The genome sequence of <i>Drosophila melanogaster</i> .";
R	L	Science 287:2185-2195 (2000).
R	N	[2]
R	P	SEQUENCE FROM N.A.
R	P	Caden E., Dreano S., Lelaure V., Mottier S., Galibert F.;
R	T	"Sequencing the distal X chromosome of <i>Drosophila melanogaster</i> .";
R	L	Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
R	N	[3]
R	P	SEQUENCE FROM N.A.
R	P	Benos P.;
R	L	Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
R	R	EMBL; AE003421; AAF45644.1; -
R	R	EMBL; AL003102; CAF19845.2; -
R	R	FlyBase; FBgn0025390; EG:5667.1.
R	R	InterPro; IPR002557; Chitin_bind_Pera.
R	R	InterPro; IPR002365; P-rich_extenam.
R	R	Pfam; PF01607; CBM_14; 2.
R	R	PRINTS; PR01217; PRICHTEXTENS.
R	R	SMART; SM00494; ChcBD2; 2.
R	SQ	SEQUENCE 1795 AA; 194464 MW; 07F10C129BD557B CRC64;
Q		Query Match 12.0%; Score 902; DB 5; Length 1795;
Q		Best Local Similarity 25.9%; Pred. No. 4,4e+1;
M		Matches 330; Conservative 163; Mismatches 537; Indels 246; Gaps
O		25 ODLSSGACRG-----GGG-YSRATNCQDVCN-----OHMECCP--58
D		124 ESFNASSGCVQHQRHDPDRPPQCQKGRPHHPDCVYRYCDKNRTQPIFCACPAGT 183
O		59 ---DKRVV-----TAELSCKG-----RCPE---SFERGRECDCAOCKKD 94
D		184 IFSPVERKCLPGDDCPSTETISDSGYTPONCEAKPECAEEGFRFPDICALYYTCRLDE 243
O		95 K-----CCPDYSFCAE-----VNHTSP-----PSKKAP--121
D		244 SGTYLQTRFKCGSNSFDLERKLCRPSEVDFCFVPVGQVYAPQYPYPAPAPPLY 303
O		122 -----PSGASQ---ITKSTTKSKPRPNCKKTXYIESELITEHVSINQESSSSSSS 173
D		304 EEDVDYTGAREEQPALKEBKLOVAAGEFKPSLVNVVLQTTILEPSTAHHKYPAAYSYP 363
O		174 SSSSTIWKIKSSKNSANRELQK---TLKVND-----203
D		364 YEISS--HHRGERAENLEKEGVRKLKLSENVIVPETPATATAATREPLANDINKQA 421
O		204 -----KKNRTKKKKPKPKPVNVDAAGSLDNGDFVTTPDTSTTQHNKYSTPKITTA 255
D		422 YKRVTYGTDKNDVTEAPEIKSPL----KGLHSENIVILPETTTT--TTTTTKPVLLTC 474
O		256 KPINRPSLPKPSD-----SKTSILTANKETVETKETTTNNKQISTDGKERT 304
D		475 PTISPDDPTPKPSSTTVTAVKSTPKISTSTOHSHTTAKTTTKRPPTVTEKTSATEKPRT 534
O		305 TSAKETQSIEKTSADIALFTSKVLKAPTEKAETTTGPAALTTPKKEPTPTPKBPASTPK 364
D		535 TVVTTTTQGRSTTTANTSDPKITTIRSTLSLKPTTTTSTTTPST-----TTTS 588
O		365 EPPTTIKSAPTTPKEBAPTTPKSAATTPEKBAPITTK-EPAATTPKEBAPTTPKEBAPT 423
D		569 TTTTPT-----TTTSTTPTSTTTVVSHNRPTTSQKTTTASTTKK-----TTSPKTT 639
O		424 TTKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEPTPTTPKEBAPTTPKEBAPTTPKEB 483

Db	640	KTIDIPSTISKSJSTTQKTTTHHKAATASTSEKKTTEKTSVSTTTKSTSESPK	699
Qy	484	PTAPKAPAPPTTPEKAPAPPTTPEKAPAPPTTKEBSPPTTPEKAPAPTTKSAPTTKEKAPPTTK	543
Db	700	PTSGTGKPTTTPKPKSTRTPTTPTKATTTTQITTTTTPLRSSSTETSTOPEPTTTTQOPPTTT	759
Qy	544	SAPPTPEKSPSTTTKEKAPAPTTPEKAPAPTTK-----KCAPTTKEKAP-----	586
Db	760	TLVTVPKSTTTTTEKPTISSPPPTTQKTTSTADPNTKVAITTOKEPTTIOSTSTTT	819
Qy	587	-----TTEKAPAPTTTTPKAPAPKAPAPPTTK-ETAPPTPKLNTPTPEKLAAPTTPEK-	640
Db	820	FTTKTTNNPPTSTEKPISTTTPEKSTTTPKSTVASTEKTTISSPKPTTEKSTENPT	879
Qy	641	-----APTTPEBLAPTTPEEBPTTPEEBAPTTTPKAABNTPEKAPAPTTPEKAPAP-	691
Db	880	TNSVKTSLTSTQRASTSTSEPKTT-QNITTTTPKPTLKISTOBATSTQKSVTIT	938
Qy	692	TPKXAPAPTTPEKAPAPTTPKGAPAPTTLKEKAPAPTTPEKAPAPKELAPTTTKEPT--STSDGK	749
Db	939	TTKATATSSPTLTTLSTEBEPTTPKPLRTPTTPTTTSVATRTKITTTTTSSESTETSTQK	998
Qy	750	APTTPKGAPAPTTPEKAPAPTTPEKAPAPTTPKGAPAPTTLKEKAPAPTTPKKAPAPKELAP--TTTK	808
Db	999	KSTTPSTTRTTPKATVIYSTQMPPTTTSKTSITVTI-----TTP-NPSETOKPTTTT	1052
Qy	809	GPTSTSDKPAAPTTPEKAPAPTTPEKAPAPTTPEKAPAPTTPEPTPEPTTBEVSTPTT-TKE--	865
Db	1053	QPTSTASTSTSIGTRIPPTTNPNQNSTSTDLTVTTRPCCPDPOSTSDKNITACTOELQ	1112
Qy	866	-----PTTHKSPDSTPELSAPPTPKALENPKXBGVTTKITPATK	908
Db	1113	QVNLELQSPQKQOFTHTRTHTLTGSRNTLGGQEVDPVDDQPSAALAESGOATTAKA	1172
Qy	909	PEMTTIAKDKTTER---DIATTPETTTAAAPMTKEATTTTEKTTESKITATTTTOVSTTT	965
Db	1173	PTMGTLAAHLLOGLFHIISTTPSKREHAP--TORPSSQBSQGR-R-GVIAQMAHNL	1229
Qy	966	QDTTPPKITTLTKT---TLAPKVTTKITITTEIMNKPEETAKPKORATNSKATTPKP	1021
Db	1230	ATSKPFIHSLRLSIQULASTQKSIDPPKILVTHNTTKEPED--SEYDSESTSEQYTEDDN	1288
Qy	1022	QKPTKA-PKKTSTSKKPKMTMRVAKPKTTPPRKMTGTMPELNTSTRAELMLQTTTRN	1080
Db	1289	EVLKTOPERAMSSTVAVALPVPSTTTERBPQK-TSSSP--SPT---KATSSITTOPI	1341
Qy	1081	QTPNSKLEVNPKSED	1096
Db	1342	ETTGGDLEHYDSGSSD	1357

DR InterPro; IPR002965; P rich extenbn.
DR PRINTS; PR01217; PRICEXTENSN.

FT NON TER 1 1
FT NON TER 1349 1349
SQ SEQUENCE 1349 AA; 135600 MW; 4DC3C1544F15EBA CRC64;

Query Match 11.9%; Score 896.5; DB 4; Length 1349;
Best Local Similarity 26.0%; Pred. No. 6.5e-41;
Matches 336; Conservative 99; Mismatches 462; Indels 393; Gaps 35;

98 PDVESCAVHNFTSPSSKAPPPGCA-----SQTIKSTKSPKP--PNKKTKK 147
112 PTTSTASASTASKTSGPTTSPVPPTTSTIAPRTSTTSASTSTTSGPTTSPVPT-- 169
146 VLESEITEHVSSENOSSSSSSSSSTI-----NK----- 181
170 -----TSTASVSKTSTSHVISISKTHSQPVTRDCHLCTWKMPDVPSPGPHGDK 222
182 -----IKSKNSANRELQKLVKDNK-----KNRTKKP 212
223 EYNNIIRSGEKICRRPEITRLQCAASHPEVSIHLAGVQCSREBGLVCRNODQGP 282
213 -----TPKP-PVDEAGSGLDNGDFVTTPTDSTQHNKVSIPKITAK 256
283 FKKCLNVEYRVLCCETPKGCPVTSTPTVTAAPSTPSGRATSTOSTSSWOKSRITTLVTST 342
257 PINPR---PGLPNSDTSKETSITVNKETTVETKETTINKQSTDG--KERTSAKQ 311
343 TSTPTSTTSAFTSTTSAFTSTTSAFTT---STSTPTSTISASAPSTSTTSAFTS 397
312 SIE-KTSAKDLAFTSKVLAKPTPKAETTKGPAITTPKEPTPTTTPKEAPTPKEPTPT 370
398 TISARTTSTISAPSTSTSP---TSTSATTTSTTSAFTSTSTPTOSTKTSASTST 454
371 TKSAPTPKEAPTTTSAFTTPK-----EPAPTTKEAPTPKEAPTTT----- 418
455 TSSSGTT---PSEVITTTSTASVSKTSTSHVSKTHSQPV--TRDCHPCTWKMPDVP 510
419 -EPAP----- 422
511 PGPFGHGDKETNNIIRSGEKICRRPEITRLQCAASHPEVSIHLAGVQCSREGL 570
423 -----TTKSAFTTPKE-----PAPTPPKAPPT----- 446
571 VCRNODQGPFKKCLNVEYRVLCCETPKGCPVTSTVTAAPSPVSGRPATOSTSSWOKS 630
447 -----TPKEAPTPKEPTTTPKEAPTPK--EPAPTTKEAPTPAKKP 490
631 RTTLVTSITSTTOSTTSAFTTSTTAPSTTSPASIPSTTSAFTTSTTSAFTTST 690
491 APTTPKEAPTPKEAPTTTKEAPSTTPKEAPPT--TSAFTTPKEAPTTTSAFTTP 549
691 TQOTTSSAPSTSTTSAFTTSTTISAPTTSTISAPTTSTTSAFTTSTTSAFTTN 750
550 KESPSTTPKEAPTPKEAPTPPKAPATTPKEAPTPKEAPTPPKAPATPKAPA 609
751 TTSAPTTSTTSAFTTSTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 810
610 PTTKETAPTPPKULTPTTPEKLAFTTPKEAPTPKEAPTTPEELAPTTPEEPPTTPEEPATTP 669
811 PTTSTTSAFTTSTTSTPTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTSTST 870
670 KAAAPTPKEAPTPTPKEAPTPTPKEAPTT----- 700
871 TSTPTSTKTSAASTSTTSGSGTTPSPVPTTSTASVSKTSTSHVSKTHSQPVTRDCH 930
701 -----PKET----- 704
931 PRCWTWKFDVDPSPGPHGDKETNNIIRSGEKICRRPEITRLQCAASHPEVSIH 990
705 -----APTPKGTAPTTLKEAPTPTPKKA 729
991 LQGVQCSREBGLVCRNODQGPFKKCLNVEYRVLCCETPKGCPVTSTPTVTAAPSTPSGRA 1050

QY 730 PK-----ELAPTTTKEPTSTSDKAPTPPKGAPTPTPKEAPTPTPKKT 781
DB 1051 TSTPTSTSSWOKSRITTLVTSTTS---TPOTSTTSAFTTSTTSTSTSTSTST 1107
QY 782 APTTLKEAPTPTPPKAPKELAPTTTSGPTSTSDKAPTPTPKEAPTPTPKEAPTPPK 841
DB 1108 APTTSTTSAFT-----HRTTSGPTTST--LAPTSTTSAFTTSTTSAFTTST 1154
QY 842 PAPTPPEPTPTTSEVSTPT--TKKEPTTIKSPDESTPELSAPTPKALNSPKPEGVP 899
DB 1155 SASITSTISAPTTSTTSTSTSTSTSTPTOSTKTSAAATSTTSGSGTTP-----SP-----VP 1205
QY 900 TTKTPAATPEMTTITAKOKTERDLRTTPELTAAPKMTKETATTEKTESKITATTQ 959
DB 1206 TTSTTSAFTTSTTSAFTTSTT-----SGGTPSPVPSTSTTSAATSTTSAFTTST 1260
QY 960 VTSTTTO--DTTPFKITTLKTTTLAPKVTYTKTITTEINMKPEETAKPKDRATNSKAT 1017
DB 1261 PTTSMTSGPTTSPVPPTTSTTS--APTSTTSGPTTSPVPPTTSTTSAFTTSTSGPS 1319
QY 1018 TPKPKPTKAPKPK--TSTKPKTMRVRKP 1046
DB 1320 TSPVPPTTSTTSAFTTSTTSAFTTSGP 1349

RESULT 11

QY 096449 PRELIMINARY; PRT; 1489 AA.
AC 096449;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Cyst germination specific acidic repeat protein.
GN CAR90.
OS Phycochlorophora infestans (Potato late blight fungus).
OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC Phycochlorophora.
OK NCBI_Taxid=4787;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=face 1-11;
RX MEDLINE=20120376; PubMed=10656583;
RA Goernhardt B., Rouhara I., Schmeltzer E.;
RT "Cyst germination proteins of the potato pathogen Phycochlorophora
infestans share homology with human mucins".
RL MOL. Plant Microbe Interact. 13:32-42(2000).
DR EMBL; AF061185; AAC72308.1; -.
DR InterPro; IPR006970; PT.
DR Pfam; PF04886; PT. 24.
SQ SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;
Query Match 11.5%; Score 863.5; DB 10; Length 1489;
Best Local Similarity 31.7%; Pred. No. 4.6e-39;
Matches 368; Conservative 55; Mismatches 509; Indels 229; Gaps 51;
QY 145 TKKYLESEITEHVSSENOSSSSSSSSSSSTTWTKSKNSANRELQKLVKDNK 204
DB 282 TPTVGYSTEEEGHGHVHTGVEPSDETEAPTEGTY--VPRETTAAPEB---DTYAP 334
QY 205 KNRTKKKTPPKPPVDEAGSGLDNGDFVTTPTDSTT-----ONNKVSTSPKITAKPI 258
DB 335 REVTVYATPEKPDVEETTYTEBETVAPTSSEINAPBERKMHYAHIEPCDTEVMYAPT 394
QY 259 NDRPGLP-----PNSDT-----SKETSLTVNKKETVETKETT---TNKQT 296
DB 395 EETVYAPTEETTYVAPTEETTYVAPTEETPYBPEETTYVTEETTYVAPTEETTYVAPTEBKTT 454
QY 297 STDGKEKTSKAKETOSIKTSKADLAFTSKVLAKTPPAEATTTKGPALTTTPKEPTPTTPK 356
DB 455 YAPTEETTYVAPTEETPYBPEETTYVAPTEETTYVAPTEETTYVAPTEETTYVAPTEETTY 511
QY 357 EPASTTPKEPTPTTITKSAFTTPKEAPPT--TTKSAFTTPKEAPPT--TPKEAPTPPKAPA 413


```

QY 324 TSKVLAKETPKAE-----TTKGAALTTPEKEPTTTPKEAPSTTPKEPTP 368
DB 161 TSHLQKRTAAEGFVPRPDVPLVPTTQPP--TTPPRPRPRPRPRPRPRPR 218
QY 369 TTISAPPTPKAPATTTKSAPTTKAPATTKEAPATTKEAPATTKEA---PTT 425
DB 219 TYL---PPTNKPPLPVITRLPRPPPPPPPPPPPPPPPPPPPPPPPPPP 268
QY 426 KSAP--TTPKEAPATTKEAPATTKEAPATTKEA---PTT--PKEAPATTKEAPATT 479
DB 269 KPLPVITRLPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 328
QY 480 KEAPATAKAPATT--PKEAPATTKEAPATTKEAPATTKEAPATTKEA---PTT 532
DB 329 RTPPRTPPKPTTTPRPAATLPPNKPDPPTTRPRPPPPPPPPPPPPPPPP 388
QY 533 -TTKEAPATTKSAPTTKEAPSTTTKEA--PTT--PKEAPATTKEAPATTKEAPATT 587
DB 389 PTNKPPLPVITRLPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 448
QY 588 T--PKEAPATTKEAPATAKAPATTKEAPATTKEA---PTTTPPEKAPATT 645
DB 449 TPLPVITRLPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 502
QY 646 BELAPTTPEEPT--PTTPEEAPATT--PKAAPNTPEKEAPATTKEAPATTKEA--- 698
DB 503 RTVTRTP--RPLTPPTKEPTTRPTTTLPPVTRITRRPPPPPPPPPPPPPP 561
QY 699 --TTPKEAPATTKEAPATTKEAPATTKEAPATTKEA---PTTKEPTSTTSKAPATT 756
DB 562 TRAPTPPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPRPR 612
QY 757 TAPTPKEAPATTKEAPATTKEAPATT--LKEAPATTKEAPATTKEA---PTTKEPTSTTS 815
DB 613 TKTTPR--PPLPPTKEPTTR--PTTTLPPSVKTRTPPPPPPPPPPPPP 655
QY 816 DKAPATT--PKEAPATTKEAPATTKEAPATTKEAPATTKEA---PTTKEPTSTTS 873
DB 656 -KP--PTTLPVTRITRATPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 708
QY 874 DESPELSAEPPLKALENSPEKEGVTTKTPAATKEATTAKOKTTERDLTTPETTA 933
DB 709 PPAVTRITRTPR-----PTLPPPKPP--TRP--PTTLPVTRITRTPPPPP 754
QY 934 APKMTKEATTEKTESKITATTQVSTTTQDT--PKKITLKTTLTAPKVTITTK 990
DB 755 -----RPPTPPTTTLPPVTRITRATPPPPPPPPPPPPPPPPPPPPPP 801
QY 991 TTTTEINMKKEBETAKPKORATNSKATTPKQ--KETKAPKKEPTSKKKTMPRVAKPXTT 1049
DB 802 --TTLPPNKP--LPVTVTRITRTPRPLPPTTRPTKEPTTTLPPVTRITRTPPP 855
QY 1050 PTPKMTSTMBELNFTSIAEAMLOTTTRPNOTENSK--LVEVNPKESEDAGAGETPHML 1108
DB 856 PT-----RPPTPPTTTLPPVTRITRATPPPPPPPPPPPPPPPPPPPP 895
QY 1109 LRPHVMEVTPDMQ--YLPRVNO 1131
DB 896 VRTTVVPPPTQRTTIVVPAPTK 919

```

RESULT 14

```

QY 08H605 PRELIMINARY; PRT; 1607 AA.
AC 08H605;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Elictor-like mating protein M81.
OS Phycophthora infestans (Potato late blight fungus).
OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC Phycophthora.
NCBI_TaxID=4787;

```

```

RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=22169232; PubMed=12180924;
RA Fadhilatus A.U., Cvitanich C., Judelson H.S.;
RT "Stage-specific gene expression during sexual development in
Phycophthora infestans."
RL Mol. Microbiol. 45:1057-1066(2002).
DR EMBL; AF507055; AAN37688.1;
SQ SEQUENCE 1607 AA; 173274 MW; 9D5A588EAD3A11C CRC64;

Query Match 11.0%; Score 829; DB 10; Length 1607;
Best Local Similarity 28.8%; Pred. No. 3.8e-37;
Matches 371; Conservative 113; Mismatches 445; Indels 360; Gaps 72;

QY 99 DYESFCAEV-----HNPSPSSKAPPPSGASQTLSKST----- 132
DB 380 DYE-YCELDENYTTIGGEWLGHSEKHPDFLWPAKPSAGVTDGISTADVALIEE 438
QY 133 -----TKRSPKPNKKTKKVVIESEITEHSVSENQSSSSSSSSSTIKIXSK 186
DB 439 SLDCVDVTSDFPDTSSTK-----PGDITKSPSYTKTPT----- 476
QY 187 NSAAARELOKKLVKDNKKNRTKKKPTKEPVVDAGSLDNGDFKVTTPDTSTQHNKV 246
DB 477 -----KSPPTKSPPTKSPPT--KPTTPPETVAPETAPATVPPEVSPETTAP 523
QY 247 STSPKITTAKPINRPSLPPNSDTSKETSLS-----TVNKETTV-----ETKETTNN 293
DB 524 STYPEPTYA--PETAPATVPPEVSPETTAPSTVPPEVAPETAPATVPPEVSPETT 582
QY 294 KQT---STDGKEKTTSA---KETQSIKTSKADLAPTSKVLAKETPKAEIT---TKGPA 343
DB 583 PSTVPEVTAETAPATVPPEVSPETTAPSTVPPEVAPETAPATVPPEVSPETT 639
QY 344 LTPK--EPPTTPKEAPSTTPKAP--TPTTKSAPTT--PKEAPTTKSAPTTKAP 398
DB 640 TTAISTVPPRRHTHPPSSNVPATVSPET--TAPSTVPEVTAETAPATVPPEVSP 697
QY 399 TTKEAPATT--PKEAPATTKEAPATTKSAPTTKEAPATT-----PKKAPATT--P 448
DB 698 ETT--APSTVPEVTAETAPET--APATVPPEVSPETTAPSTVPPEVTAETAPATVP 751
QY 449 KE--PAPTPKEBETPTT--PKEAPATTKEAPATTKEAPATT-----APKAPATT-- 494
DB 752 PETVSPETTAPSTVPPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAPATVP 811
QY 495 -----PKEAPATT--PKEAPATTKEAPSTTPKAPATTKSAPTTKEAPATTKSAPT 547
DB 812 PETVSPETTAPSTVPPEVTAETAPET--APATVPPEVSPETTAPSTVPPEVTAETAPATVP 870
QY 548 TPKEPSPTTKEAPATT-----PKEAPATT-----PKKAPATT--PKEAP--TTPK 590
DB 871 PPEVSPETT--APSTVPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAPATVP 927
QY 591 EPAPTTTKKAPATKAPAPATT--PKETAPTT-----PKLAPTT-----PEKLA 634
DB 928 ATVPPEVSPETTAPSTVPPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAP 987
QY 635 TT-----PEKAPATT-----PEELAPTT-----PEEPPTT-----PEEBAP 666
DB 988 ATVPPEVSPETTAPSTVPPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAP 1047
QY 667 TT-----PKAAPNT-----PKEAPATT-----PKEAPATT-----PKKAP 698
DB 1048 ATVPPEVSPETTAPSTVPPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAP 1107
QY 699 TT-----PKETAPTT-----PKGAPATT-----EPAPTTKAPAPKELAPTTK 739
DB 1108 ATVPPEVSPETTAPSTVPPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAP 1167
QY 740 ---EPSTTSKAPATT-----PKGAPATT-----PKEAPATT-----PKKAP 775
DB 1168 ATVPPEVSPETTAPSTVPPEVTAETAPATVPPEVSPETTAPSTVPPEVTAETAP 1227

```

```

Qy 776 TT-----PKGAPTT-----LKBPATTPKKAPKELAPTT-----TKGPTSTS 815
Db 1228 ATTPPETYSBETTABSTYPPETVAPETTABATTPPETYSBETTABSTYPPETVAPETTAB 1287
Qy 816 DKPAPTT--PKETAPTT-----PKBPATTT-----PKKAPTT--PET--PPPTTS 856
Db 1288 ATTPPETYSBETTABSTYPPETVAPETTABATTPPETYSBETTABSTYPPETVAPETTAB 1347
Qy 857 VSPPTTKERTTIHKS---PDESTPELSAEPKALENSPKBGVPTTKTPAATKEMT 913
Db 1348 ATTPPETYSBETTABSTYPPETVAPETTABATTPPETYSBETTABSTYPPETVAPETTAB 1406
Qy 914 TAKDKTERDLRTTPETT-----TAAPKMTKEATTEKTTESITATTTQVSTTT 965
Db 1407 PA-----TTPPETYSBETTABSTYPPETVAPETTABSTYPPETYSBETTABSTYPPETVAPETTAB 1461
Qy 966 QDTTPPKITTLTKTTLAPKVTTKTIT-----TTTEIMNKBEETAKPKDRAATNSKAT- 1017
Db 1462 PETT--APATYPPETYSBETTABSTYPPETVAPETTABATTPPETYSBETTABSTYPPETVAPETTAB 1518
Qy 1018 TTPPKAPPTAKPKPTSTTKKMTKPKTKTTPPKMTSTMBELNFTSKIAEAMLOTTT 1077
Db 1519 TPTTAPSDSPNYPTETPESTDAP--TYAPTSTPTETSVVYAPETVAPETPTIV-----TY 1570
Qy 1078 RPNQTPNSKLVBNPKSEBAGAGETPH 1106
Db 1571 KPTSTPKS-----TNP-----GOANOSPN 1589

RESULT 15
Q9SPMO PRELIMINARY; PRT; 1315 AA.
AC Q9SPMO,
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Extensin-like protein.
GN PEX2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OC NCBI_Taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen;
RA MEDLINE=21310248; PubMed=11437249;
RA Stratford S., Barne W., Hohorst D.L., Sager J.G., Cotter R.,
RA Golubiewski A., Showalter A.M., McCormick S., Bedinger P.;
RT "A leucine-rich repeat region is conserved in pollen extensin-like
RT (Pex) proteins in monocots and dicots.";
RL Plant Mol. Biol. 46:43-56(2001).
DR EMBL; AF159297; AAD5980.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007090; LRR.plant.
DR InterPro; IPR002965; P-rich_extensn.
DR InterPro; IPR001998; Xylose_isom.
DR Pfam; PF00560; LRR_3.
DR PRINTS; PR01217; PRICEXTENSN.
DR PROSITE; PS05052; LRR_PS_2.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1;
SO SEQUENCE 1315 AA, 134502 MW, 64C97A2A01F0936F CRC64;

Query Match 10.8%; Score 816; DB 10; Length 1315;
Best Local Similarity 23.5%; Pred. No. 1.6e-16;
Matches 265; Conservative 106; Mismatches 486; Indels 270; Gaps 35;

Qy 73 RCESTFERGECDDQCKYDKCCPDYBFCABVNNPTSPSSK-KAPPPSASQTITS 131
Db 390 QCAPVLARPEVDCS-----KAVCAGYPTPGGPSSSVGKRPSPGK----- 432
Qy 132 TTGSPKPPNKKTKVIESEITEHVSSENQSSSSSSSSSSSTIWKIKSKNSAAN 191

```

```

Db 433 --PAAPAPRTHTTPDVSEPLPESPV-----PAAPARMPTLRSPRD 476
Qy 192 RELQKKLVKNNKNNKTKKKKPTKPPVVDAGSLDNGDKVTTPTDSTQHNKYSTSK 251
Db 477 EYI-----PTPEVPKSSPPG-----TSPASHGA--PLQAQPP 508
Qy 252 ITTAKRINRSLPNSDTSKETSLSLVNKEITVETKTTTKQISTDCKETSAKQ 311
Db 509 AASSPPATVKSPPPAV----- 527
Qy 312 SIEKTSADLAFTSKVLAKPTKAEITTKGPAITTPKEPTTPPKEPASTPKPTTI 371
Db 528 -VLPPAKTPSPPAV-ASPPPEAVSSQPVKSSPPPAVASPPPKSP--PPPAV 583
Qy 372 KAPPTPKPAATTKSAFTT--KEBP-----TTKEPAPTPKPAATTKKAPAT 423
Db 584 ASPPPLMKSPPPPAVASPPQPLKSPPVLMISTPSVKSPPPPVVASPPPVKSPPL 643
Qy 424 TTKSAPTTKE-----PAV--TTP-----KKBPATTKBPATTPKE--PTPTTKBPATTK 472
Db 644 AVSSSPSPVKLPPLAPGKSTPPEEEKPTPTPVKSSPPPEKSLPPTLTSPPOK 703
Qy 473 BPAPPTPKPAATPKAPATTPPKBPATTPKEBPATTKKBSPTTPKBPATTKSAFT 532
Db 704 PTPPTSPKSPSPSPVETLPPPSKSPPEEVSSPPQAKSSP-----PAVVS--SPRP 756
Qy 533 TTKBPATTKKAPATTPKBSPTTPKEBPATTPKEBPATTPKKAATTPKEBPATTPKBP 592
Db 757 LKSSPPPVSSPPPTPKSPPPLAVSSPPQVKTSPPPAVSSPPPTPKSPPPLAVS 816
Qy 593 APPTTKKPAATPKBPATTPKETAATTPPKLPTTPEKLAATTPKEAPTTPELAFT 652
Db 817 PPQVKTSPPPAVSSPPPTPKSPPPLAVSSPPQVKTSPPPAVSSPPPLBKSPPS 876
Qy 653 PPEPTTPPEBPATTPKKAANTPKBPATTPKBPATTPKBPATTPKETAATTPPKCT 712
Db 877 SVSSPTTVKSSPPPAVSSPPMTKSSPPPAHVSSPPAEKSSPPLAISSPPSPSPK 936
Qy 713 A-----PTTKBPATTPKKAPKELAPTTKPTSTSDKAPT-----TPKG 756
Db 937 SSPPVKTSPPPAVSSPPPTPKSPP--APVSSPPPVKSSPPPAVSSPPPTPK 993
Qy 757 TAPTP-----KEBPATTKBPATTPK--GAPTTLKBPATTPKKAAPKELAPTT 806
Db 994 LPPEAVSSPPPVVSSPPPTVSSPPPTKLPPEPTVSSPPPTXKLP--APVS 1050
Qy 807 TKGPTSTSDKAPATTT--PKETAPTTPK-----EPATTPKKAAPTTETPTTSEVSTP 860
Db 1051 SPPPVKSSPPPAVSLPPTTKKSPPRKRVSSPPVVKCCPPTLVSSPPPAKSLPPT 1110
Qy 861 TTKKPTTIHKSPEDESTPELSAEPKALENSPKBGVPTTKTPAATKEMTTAKDKTT 920
Db 1111 TVVSSP-----PPEVKSSPPPTPVSSPPPAKSSPPPTPVSSPP----- 1150
Qy 921 EEDLFTT-----ETTAAPKMTKETATTTKTSKITATTTQVSTTTQDTPPKITT 975
Db 1151 --ELKSSPPPAVSSPPSAKSSPPPAVSLPPEVKSSPPPAVSSPPPAKSP----- 1203
Qy 976 LKTTLAPKVTTKTITTEIMNKBEETAKPKDRAATNSKATTPKQKTKAPKPTSTK 1035
Db 1204 -----PPPA-----PMSLPPPVKSSPPPAVSSPPPMKSPPPPAVPS 1242
Qy 1036 KPKTPRAKPKPTTPPKMTSTMBELNFTSKIAEAMLOTTTRPNQTPNSKLVBNPKSE 1095
Db 1243 SPPPAV--VKPSSLPPPAVSSPPAV-----TSAP-----PKKE 1275
Qy 1096 DAGGAGETPHHLARPHFMPEVTDPMDYLPVNPQGIINMMD 1142
Db 1276 EDSTA-----PPAEALPPSSFNDIILPEIMANK 1303

```

Mon Dec 8 09:50:45 2003

us-09-556-246-1.rpt

Page 16

Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:33:27 ; Search time 27 Seconds

(without alignments)
5000.771 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526
Sequence: 1 MAMKTLPIYLLLLSVFVIQ.....AAATTRSGQTLSKWWYVNC 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616882 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1090	14.5	3020	2 A43932	mucin 2 precursor,
2	950	12.6	1664	2 T18262	S-layer protein -
3	863.5	11.5	1489	2 T31108	cyat germination s
4	846	11.2	1274	2 T16251	hypothetical prote
5	832	10.8	2187	2 T30826	nascen polypeptid
6	809	10.7	1367	1 S48478	glucan 1,4-alpha-g
7	784	10.4	1188	2 S49915	extensin-like proc
8	695	9.2	1151	2 T18535	high molecular mas
9	676.5	9.0	1229	2 T25697	hypothetical prote
10	672	8.9	1344	1 A35175	mucin 1 precursor,
11	666.5	8.9	7962	2 T38346	elastin ctfn - hu
12	663	8.8	3570	2 T45025	mucin MUC5B, trach
13	658.5	8.7	3507	2 T34513	hypothetical prote
14	643.5	8.6	990	2 T51618	nucleolar phosphop
15	633	8.4	489	2 T11622	extensin class 1 p
16	632	8.4	761	2 C84672	hypothetical prote
17	629	8.4	6642	2 T29757	protein UNC-89 - C
18	627.5	8.3	971	2 T19431	hypothetical prote
19	622.5	8.3	839	2 T75518	hypothetical prote
20	607.5	8.1	801	2 T29018	hypothetical prote
21	607.5	8.1	924	2 S27923	gene Lrf protein -
22	605	8.0	379	2 S50125	larval glue protei
23	594	7.9	4776	2 E95206	cell wall surface
24	591.5	7.9	2232	2 T34434	hypothetical prote
25	586.5	7.6	1630	2 A53577	secreted sialoglyco
26	586	7.5	883	2 E90577	11p0protein vacuol
27	563	7.5	880	2 D89756	protein T32E7.2b l
28	559.5	7.4	350	2 S22456	hydroxyproline-ric
29	559	7.4	875	2 S23760	polyphenolic adhes

30	556.5	7.4	856	2 T16543	hypothetical prote
31	551	7.3	620	2 S06733	hydroxyproline-ric
32	551	7.3	1087	1 QFMSH	neurofilament trip
33	545	7.2	873	2 A47283	calphostin - fruit
34	542	7.2	369	2 S20500	hydroxyproline-ric
35	540	7.2	1459	2 T33271	hypothetical prote
36	538.5	7.2	416	2 J00465	extensin precursor
37	537	7.1	1072	1 A37221	neurofilament trip
38	532	7.1	756	2 T27642	hypothetical prote
39	530.5	7.0	813	2 S70795	vsaa protein precu
40	530.5	7.0	1162	2 JH0557	exo-alpha-sialidas
41	530	7.0	865	2 A47282	calcium-binding pr
42	522	6.9	328	2 J00985	hydroxyproline-ric
43	521.5	6.9	854	2 S02003	neurofilament trip
44	518	6.9	866	2 T45462	membrane glycoprot
45	518	6.9	1611	2 T38236	hypothetical prote

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (Fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #ext change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
J:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Klm, Y.S.
J: Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A:Reference number: A49963; MUID:94132002; PMID:8300571
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: GB:121998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Klm, Y.S.
J: Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upst
A:Reference number: A45106; MUID:93016075; PMID:1400449
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:9186397; PIDN:AAA59164.1; PID:9186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.J
J: Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymor
A:Reference number: A43932; MUID:91358717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Klm, Y.S.
J: Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evid
A:Reference number: A33533; MUID:89197956; PMID:2703501
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:9188873; PIDN:AAA6334.1; PID:9188874
A:Experimental source: intestine
R:Tany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Klm, Y.S.; Bebbum, C.B.

J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; PMID:91086481; PMID:1985113
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R:Xu, G.; Han, L.; Khatir, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-T
A:Reference number: PQ0328; PMID:92198477; PMID:1550588
A:Accession: PQ0328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M6523
A:Experimental source: small intestine
A:Accession: PQ0329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
A:Genetics:
C:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: 11p15.5-11p15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C:Keywords: glycoprotein; intestine; tandem repeat
C:Keywords/Domain: von Willebrand factor type C repeat homology <WMC>
'2766-2834

Query Match	14.5%;	Score 1090;	DB 2;	Length 3020;
Best Local Similarity	27.4%;	Pred. No. 2.4e-37;		
Matches 378;	Conservative 107;	Mismatches 514;	Indels 382;	Gaps 51;

```

QY 43 TCNCDY--NCQHYMECCDPFRVCTALBSCRGRCFSEFSEGRECCDCAQCCKYDKCPDY 100
Db 1079 SCSCDDTGDDGDCFCFSAVASVYAQECCTKEGAC-----VPMRPPDL-CP1FCDDYLN---PPh 1128
QY 101 ESFCAEHNPNFSPSSSKAPPPSGASQTI-----KSTTKSEPK--PPNKKTKKVIASEE 153
Db 1129 E--CEMHHEPCGNRSFETCRTINGIHSN1SVSYLEGCPRCPCORPIYEBDLKCVTAJD 1186
QY 154 -----ITEEH-----SVSENOESSSSSSSSST1WKIKSNGSNAABELQKJLKVKCNKK 205
Db 1187 CGCVCVEDTHYPGASVPTFEHRCVKCVCTNSQVCRPEEGK-----IL 1229
QY 206 NRTKKKPLPKPPVVDAGSLDNGDF--KATTPDSTTQH--NKVSTSKITTAAPIN 259
Db 1230 NOTQ-----DGAFCWYELCGBNGVKEHFN1CSITTRPSTLTFTT1 1271
QY 260 PRPSLPNSDTSKETS1VNKETTVETKET--TTTNK-----QSTQKE 302
Db 1272 TLPPTP-----TSFTTTTTTTTTSS1VLS1TTPK1LCLMSW1MEDHSSGSDGDR 1323
QY 303 KTTTS--AKETOS1EKTSAKD----- 320
Db 1324 EPPFGVCGAPEDICRSVQDPHLSLEQHQAQVOCDSVGFICKNEDQFGNGPGLCYDYK 1383
QY 321 -----LAPTSKVLAKPPTKPAETTKGALTTTPKSPPTTPKREPASTTPKSPPTP 368
Db 1384 IRVNCCEPMDKCIPTTSPPTTTPSPPTTTTLPPTTTPSPPTTTTTPPTTTPSPPT 1443
QY 369 TTINSAP--TTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTTS 427
Db 1444 TTTTPPLPTTPSP1S1TTPPTTPPTTSP--PTTTPSPPTTTPSPPT--TTTTTPPTTPPS 1501
QY 428 AP--TTPKEBAPTTPKKAAPTTPKEBAPTTPKEAPTTPTPKKEBAPTTPKEBAPTA 486
Db 1502 PPMTPPTTPPASTTLPPTTPPSPTTTTTPPTTPPSPTTTPP1TP--PTSTTLPPTT 1560
QY 487 PKKAPAPTTPKEBAPTTPKEBAPTTPKBSPTTPKEBAPTTPKSAAP--TTTKEBAPTTPKSA 545
Db 1561 TTPSPPTTTTTTPPTTPSPPTTTPSPPTTTPSPPTTTPPTTPPTTPPSPTTTTTPPTTPSP 1620
QY 546 P-TTPKBSPTTPKKEBAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKKEBAPTTPKKAAPTA 604

```

Db	1621	PTTPIRPIRPISTTLRPTTTSPPPTTTTTPTTTSPPPTTTSPPPTTTTTTTPPTTT	1680
Oy	605	PKBPAPTTPKETAAPTTPKULPTTPEKLAPTTPEKAPTAETPEELAPTTPEEPPTTPEEP	664
Db	1681	PSSDITTTT--SPPTT--TMTTTS--PTTTPSSPITTTTTTSPSTTTP--SPPTTMTPT	1731
Oy	665	APTTPKAAAPTPEKBPAPTTPEKBPAPTTPEKBPATTP-----KETAPTT-----	709
Db	1732	SPPTTSPPTTMTTLRPTTTSSEPLTTPPLPSSITPTTSPSPSTTPTTTPCVPLCMWTGW	1791
Oy	710	--KG-----	711
Db	1792	LDGKPNFHKGGGTIDGVCGGMAINSCTATMPDVIFIGQLGQTVCDVGLICK	1851
Oy	712	-----TAPTTLKEBAPTTPKKPAPKEAPTT--TKBP	741
Db	1852	NEDQKPGVIMAPCLANWEINQCCQCTQTMT--TTTTTMTPTTPIPTTTTIVPTP	1908
Oy	742	TSTTSDKP---APT--TPKGTAPTTPEKBPATTTPEKBPATTPKGTAPTTLKEBPATTPK	796
Db	1809	TPITSGNGLQAPTPPISITTTTPTTPTPTGTQTPPL--TPIPTTTTIVPTTPTPTGTQ	1967
Oy	797	PAPKEAPTTKGTST--TSDKAPTTPEKAPTTPEKBPATTPKKP-----APTPE	848
Db	1868	PTTVLITTTTTMTPTTPTSTKSTTVPIPTTTTIVAPTPPTGTQTPMTAPISITTTVTP	2027
Oy	849	TPPTTSVSTPTTTK-----EPTTIKSPDESPTPELSAETPRALENSPKBPVPTTKT	903
Db	2028	TPPTTSGTPPTTITSTAPLAEITLNSPPESSITPQISRSITSSPLTESITLSTLPAIE	2087
Oy	904	PAATKPEMTTADKTERDLRTTPETTTAPAKMTKEATITTEKTESKITATTTQVYST	963
Db	2088	MTSTAPSTTPAPTTTSGCHTLPSPBSTTSSPPTPRGTTT--GSSAPPTSTVQTTT	2145
Oy	964	TTQOTTPPKITT--LKTITLAPKVTITTKITITTELMNKEP-----	1003
Db	2146	SAMTPTPLPSTPSAIRITGTLRPSPSVLICVILNDIYVAGEEYVNGTYGDTCYFVNC	2205
Oy	1004	-----TAKPKDRATNSKATPKPOKPTPAKPKPTSTKPKPTMPVRKPKTTPT	1051
Db	2206	LISCTLEFYNNMSCPTSPPTPSK--STPTPSKPSSTSKPRPKTPRECDFDPR-----	2260
Oy	1052	PKMTSTMPBELNPTSRIAEAMLTQTTTPNQTPNKSLVNP-----KSEDA	1098
Db	2261	-----QENETMMVLCDCFM--ATCKXNNTVEIYKVECEBPMPWPTCSNGLQAPRAVED	2310
Oy	1099	G-----ABGETPHML-----RPHVMEVTPDMVDYLRVNVOGITII	1135
Db	2311	GCCWHMECDYCTGMD--PHVTFEDGLYYSQNGCTVLVLEISPSVD-----NFGYI	2363
Oy	1136	N 1136	
Db	2364	D 2364	

RESULT 2

S-layer protein - *Clostridium thermocellum*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

R; Fujino, T.; Beguin, P.; Aubert, J. P.

A;Title: Organization of a *Clostridium thermocellum* gene cluster encoding the cellululosoma

A;Reference number: Z18847; MUID:93209931; PMID:8458832
 B;Accession: T18262

A;STATUS: preliminary
A;ACCESSION: 116462

A;Molecule type: DNA
A:Residues: 1-1664 <FUJ>

A;Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

QY 785 -----TLKEPAPTPPKP-----APKE---LAPT--TTGPTSTSDKAP 820
 Db 1048 ASTEETVAPTEETVAPAEETVPEETETVAPTEETVAPTEETVAPTEETVAPTE 1107
 QY 821 TTPKETAPTPKEAPTPPKAPPTTETPTTSEVSTPTTKPT--TIHKSDESTP 878
 Db 1108 ETVVAPAEETVPEETETVAPTEETVAPTEETVAPTEETVAPTEETVAPTEETV 1166
 QY 879 ELASAPTEKA--LENSPKRG-----VPTTKPAKPKETTTAKKTERDRLRTP-- 928
 Db 1167 YAPTEETVAPTEETVAPTEETVAPTEETVAPTEETVAPTEETVAPTEETVAPTE 1226
 QY 929 ETTTAPAPTKETATTTETKTESKITATTQVSTTQDPT--PFKITTT-- 975
 Db 1227 BETTAP-----TEETVPEETETVAPTEETVAPTEETVAPTEETVAPTEETVAP 1281
 QY 976 LKTTTLAPKVTTKTITTTETIMNKPEETAKPKDRAT--NSKATTPKP--QKETA PKKPT 1032
 Db 1282 TEETVAPTEATVAPTEETVAPTEETVAPTEETVAPTEETVAPTEETVAPTEETV 1341
 QY 1033 -----STKKP--KTMKRVAKPKTTPPRKMTSTMPELANTSKIAEMLQTTTRP 1079
 Db 1342 PYPEAESSTVSTVTEKPCNTEETETDEPTDEPTDE--PSDEPTDEPTDEPTDEPT 1399
 QY 1080 --NOTPNSKLVEVNPKESEDAG 1098
 Db 1400 CDNOINGIGVENKRYNNAG 1420

RESULT 4
 116251
 hypochloral protein F35A5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T16251
 R:Leimbach, D.
 submitted to the EMBL Data Library, January 1996
 A:Description: The sequence of C. elegans cosmid F35A5.
 A:Reference number: Z18485
 A:Accession: T16251
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: EMBL:U46675; NID:q1166613; PID:q1166621; PIDN:AA52641.1; GSPDB:GNOC
 C:Genetic8:
 A:Gene: CESP:F35A5.1
 A:Map position: X
 A:Introns: 1272/2

Query Match 11.2%; Score 846; DB 2; Length 1274;
 Best Local Similarity 27.6%; Pred. No. 1e-27; Indels 248; Gaps 60;
 Matches 313; Conservative 119; Mismatches 455;
 QY 114 PSSKAPPSGASQTIKSTTRSP-----KPP---NKKTKKVISEETIEHSVENQES 166
 Db 185 PSPKKAASKMDPIVPTPIKNPAAKMKKPMWEDDEVTEETKEEPATKRVPAKKEP 244
 QY 167 SSSSSSSSSSTIWKIKSSKNSAANRELQKKLKYDNK--KQRTYK--KP-----TP----- 214
 Db 245 STSVAPVDPSPPTKV-----PVKKEPEVPPPTIKNPTKWKKPMWEDDEVTEEVK 294
 QY 215 KPPIVDEKSGLDND-----FKVTPPDISTQHNKVSISPKITTAKEP--NRPISLPPSD 269
 Db 295 EPPVPEKAPVLLKKDDPAAPAKARDPSKAPKVEPSSVVPPTPVKNPVKKYKPEWE 354
 QY 270 TSKETSILVNKETVETKETTTNNKOTSDGKEKTSKAKETOSIEKTSADLAPTSKYL 329
 Db 355 VDDEPAEAVKPSAPKEKTPVLAKKEPEPSSSTSSDPSKKAAPAVPRDSSPKKATPL 414
 QY 330 KPTPKAE-----TTTKGPA-----LTTKPEPTPTT-----KEPASTPK 364
 Db 415 QADPKAQEVPTPVKNPVKKYKPVWEVDEDDPVEEVKQPEAPAKKTPVLAKKEPAKQTA 474

QY 365 EP-----TP--TIKSAPTPKAPPTTKSAPTTKAPPTTKBPA----- 405
 Db 475 KPAISKTEPEKDPVPRDSSPKVAKDSQAP--TVQNVKVKMRPMEDETPA 533
 QY 406 -----PTPKAPPTTKAPPT-----TTKSAPTPKSP-----APTPPKAPPTP 448
 Db 534 DVSKPTDAKTPISLAKODAPAKESLKPADTAPAKARDPSPKVAPPTAPEKKTPLA 593
 QY 449 KE---PAPTPKEPTPTPKAPPTTKAPPTTK--BPATAPKAPAP-----TTKE 497
 Db 594 KKEPAGPADSKTEKESKPRDPSKAVPAPKPVKEVAAVAKKEPISKPDATAPK 653
 QY 498 PAPTPKEAPPTTKBP-----SPTPKAPPTTKSAPTTKAPPTTKBAPTTP--KEP 552
 Db 654 AEPNSPVVP--PTPVKNPVKKKPMWEDDAPAKVSLDEPEKK--TVLAKAPTKPDSEA 711
 QY 553 SPTTKAPAPPTPK--BPAPTPPKAPAP--TTPKAPPTTKAPPTTKAPPTA--PKE 607
 Db 712 AADVSGSSGSKOPKLAKAPVPRDPSPMKAVPIKPAKT--EVPRAVKKDEPVAKSRD 769
 QY 608 PAPTPKETAPTPKLTPTTP-----EKLAPTPEKAPPTPBEAPTTPEEP-----T 657
 Db 770 PSPKKAK--AEPNSP--VVPPTPVKNPVKKKPMWEDDAPAPVNVPEPEKKTPLAKKT 826
 QY 658 PTPPEBPAPPTPKAAAPVTPKEAPPTTKBAPT--PKEAPPTTKETAPTPKGTAPTT 716
 Db 827 PVKPRDPSPKAVAPKASTKTDAPVSKPEPSSKPEBSPKAEPSVVP---PTP 882
 QY 717 LKEPAPTPPKAPKAPLAPT--TTKEPTSTSDKAPPTPKGTAPPTPKAPPTPKAP 775
 Db 883 VKNVKKK--KPPWEDDETEEVKPE--PEKTPVLAK--KEDEKPRD--APVAAKPRD 937
 QY 776 TTPKGTAPTTKEBAPT-----TPKKAP--KEAPTTTKGPTS-----TSDK 817
 Db 938 PSPKKAAPVE--KEPAAVAAKPRDLSPKAIPIPANTQEAPEPTPVKNPVKKKPMWEDDE 995
 QY 818 P-----APTPKEP-----APTPKEBAPTTPKPAAPTPTEPTTSEVSTPTTKEP 866
 Db 996 PAEPVSAPEPEKTPVLAKKAPAKPRD--SPKAAAPVAAK--DDPKIPEV--PTPVKNP 1050
 QY 867 TTIHKS-----DESTEPISA--EP---TPKALENSPKRGV-----PTTKPA 905
 Db 1051 VKMKKPMWEDDESEPVSAPEPEKTPVLAKKAPTPATPQDSEAAADPVSGTSDPK 1110
 QY 906 AT-----KPEMTTAAOKTTERDLRTTPTT--TAAPKMTKETATTTETKESKIRATT 957
 Db 1111 LSKAPVPEKPRPTDPKDDKLPSPAKKPEKAPAPAKKMPWDDDPDEPEADFTVPA 1170
 QY 958 TQVSTTQDTPTPKITTLAKTTTLAPKVTTKTITTTETIMNKPEETAKPKDRATNSKAT 1017
 Db 1171 PSKAPTEEDPADPAG-----GPKTKDPK-----LKKAPAPKEPTEK----- 1206
 QY 1018 TPKEQKPTKAPKKEPTSKKPTKMPRVKPK-----TTPTRKMTSTMP 1060
 Db 1207 -PKPEVSKPEPKTEPPKP--AAPKMKPMWEDDPDEPADFTMPAPKKEPTEDP 1259

RESULT 5
 130826
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-NAC protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
 C:Accession: T30826
 R:Yoclov, W.V.; St-Arnaud, R.
 A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle-
 A:Reference number: Z20869; M0UD:96312450; PMID:8658236
 A:Accession: T30826
 A:Molecule type: DNA
 A:Residues: 1-2187 <YOT>

A.Citrob+references: EMBL:U48363; NID:G1666688; PID:G1666689; PIDN:AA18732.1
C.Genetics:
A.Gene: Naca
A.Map position: 10
A.Littrons: 24/1, 1996/1, 2024/3, 2050/3, 2099/3, 2142/3, 2183/3
A.Note: differential splicing convert aShanNC into a tissue-specific DNA-binding active
C.Keywords: alternative splicing; DNA binding; transcription factor

Query Match 10.8%; Score 812; DB 2; Length 2187;
Best Local Similarity 27.4%; Pred. No. 4.3e-26;
Matches 311; Conservative 128; Mismatches 449; Indels 246; Gaps 51;

[illegible][illegible]

RESULT 6
 548478
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
 N/Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c
 C/Species: *Saccharomyces cerevisiae*
 C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999
 C/Accession: S48478; A26877; S27281; JG6123
 R/Bowley, K.
 submitted to the EMBL Data Library, October 1994
 A/Reference number: S48478
 A/Accession: S48478
 A/Molecule type: DNA
 A/Residues: 1-1367 <ROW>
 A/Cross-references: GB:247047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN00009; MIMF:
 R/Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A/Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
 A/Reference number: A91631; MUID:87194600; PMID:3106330
 A/Accession: A26877
 A/Molecule type: DNA
 A/Residues: 1-242 <YAM>
 A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
 A/Accession: B26877
 A/Molecule type: DNA
 A/Residues: 762-1331 <YA2>
 A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
 R/Pardo, J.M.; Ibanez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Sacchar*
 A/Reference number: S27281; MUID:89031230; PMID:3141213
 A/Accession: S27281
 A/Molecule type: DNA
 A/Residues: 1-31 <PAR>
 A/Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R/Lambrecht, M.G.; Bauer, F.P.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A/Title: Muc1, a mucin-like protein that is regulated by Msn10, is critical for pseudohy
 A/Reference number: JG6123; MUID:96323237; PMID:8710886
 A/Accession: JG6123
 A/Status: nucleic acid sequence not shown
 A/Molecule type: DNA
 A/Residues: 1-1367 <LAM>
 A/Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
 C/Genetics:
 A/Gene: SGD:MUC1; STA2; MAL5; DEK2; SGD:S0001458
 A/Cross-references: MIPS:YIR019c; SGD:S0001458
 A/Map position: 9R
 C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C/Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F/5-21/Domain: transmembrane #status predicted <TM1>
 F/1350-1366/Domain: transmembrane #status predicted <TM2>
 Query Match 10.7%; Score 809; DB 1; Length 1367;
 Best Local Similarity 27.5%; Pred. No.3.6e-26;
 Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53;
 85 DCDAGCKKDKD---CCPDYESCFAEYHNPTSPSSKAPPGSGASQSTIKTSRKPFPN 141

```

Db      174 DLSTGNNVYDNGHSGTDFPGFYWNIDCDNNGCGKSSSTTSSSES--STT----- 223
Qy      142 KKKTKKVESEITEHEHVSVENQSSSSSSSSSSSTTWWKSKSNKSNANNELOKKLVK 201
Db      224 ---TSSTSSSTTSSSTSSSTTSSSTSSSTSSSTT-----APATPTTSSCTEKEPTP 274
Qy      202 DNKKRRTKKKPPKPPVVDAGSGLDNGDFKVTPTDSTTQHNKYSTPKITTAKPINR 261
Db      275 PTTTSSCTEKEPT--PPHND-----TTCCT-----KKKTTTSCCTCKKTTTTPV 314
Qy      262 PSLPNSDTSKETSILTANKETVEKETTNNKQSTDGKEKTSKAKETOSIEKTSADL 321
Db      315 PT--SSSSTSSS-----APVTPSSSTT-----SSSSAPVTSSTSSSES--- 353
Qy      322 APTSKVLAKPTPKAATTGKPLATTPKEPTPTTPKEPASTTPKEPTPTTISAPPTPEP 381
Db      354 APV-----PTPSSSTSSSAPVTSSTSSSAPVTSSTSSSAPVPTPSSSTSSS 407
Qy      382 APTT-----TKSAP-----TTPKEPAPTTTKEPAPTTKEPAP-----TTPKEPAPTTT 428
Db      408 APVTSSTSSSAPVTSSTSSSAPVTS-----TTESSSAPVTSSTSSSAPVPTPSS 463
Qy      429 PTPKEPAPTTPKAPATTTPKEPAPTTKEPAPTT-----TTPKEPAPTTKEPAPTTKEPAP 485
Db      464 STTESSSAPVT-----SSTTESSSAPV-----PTPSSSTTESSSAPVT-----SSTTESSSAPV 512
Qy      486 APKAPAPTTPKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 545
Db      513 -----PTPSSSTTESSSAPAPVPS-----SSTTESSSAPVTS-----TTESSSAPVPTPSS 559
Qy      546 PTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 604
Db      560 STTESSSTPTVTS-----TTESSSAPVPTPSSSTTSSSAPVPT-----DSSSTTESSSAPA 611
Qy      605 PKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPELAPLT--TPEEPPTTTEE 663
Db      612 PTPSSSTTESSSAPVT-----SSTTESSSAP--VPTPSSSTTESSSAPVPTPSSSTTESSSA 666
Qy      664 PAP-----TTPKAAPNTPKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 717
Db      667 PVPPTSSSTTESSSAPVT-----SSTTESSSAPVT-----SSTTESSSAPVPTPSSSTTESS 718
Qy      718 KEAPATTTPKKAPKELAPTTTKEPSTSDKAPTTTSGTAPTTKEPAP--TTPKEPA 774
Db      719 SAPVPTSSSTTESSSAPVPTPS--SSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSA 777
Qy      775 PT--TPKGTAPTTLKEPAPTTPKKAPKELAPTTTGTPTSTSDKAPTTTKEPAPTTTKE 833
Db      778 PVPPTSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSSNITSSAPSSPTPSSSTTESSSV 837
Qy      834 PAP-----TTPKKAAPTTPETPTTSEVSTPT-----TTPKEPTTIHKSPTDESPTPELSAE 883
Db      838 PVPPTSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSSNITSSAPSSI---PPTSSSTTESPT 894
Qy      884 PTPKALESPKEGVPPTTKTPATKPEMTTAKDKTTERDLATTPETTTAAPKOKTERAT 943
Db      885 GT--TTPSSSKTPGSGTETSSTTETTVTPKTTTSTVTPSTTTTITTVCSGTNSAGE 953
Qy      944 TTEKTESKITAT--TTOVSTTTTODTTPPFKITTLTKTTTLAPKVT--TKKTTT--EI 997
Db      954 TTSGSPRTVTTTPTTTTSTTITTTTSTTITTVCSGTNSAGEITTSOCSPTITTTVPC 1013
Qy      998 MNKPEETLAKPKORATNSKATTPKPKP-----TKAPKPTSTYKPKPTM 1040
Db      1014 STSSEITA-----SSTTTSPTTPTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTV 1065
Qy      1041 PRVRKKPTTPPRKMTSTMPBELNPTSRIAEMLQTTTPNQTNPNSKLVAVNPKSDAGDA 1100
Db      1066 PTTYLTITTAAPT--SVTTVNTFTPTT--ITTVCSGT-----NSAGTTSGC 1109
Qy      1101 EGEPHMLLRHVMEVETPMDVLRPVNOCIIINPMLSDETNINCNGKPVGLTTLRNG 1160

```

```

Db      1110 SPKTVTTTVCSTGCTGEYTEATLTVTAATTVTVTVTSSSTGINSAG--GKTTGYTKSV 1168
Qy      1161 TLVAFRGHYFMWLSPPSPSPARITBVGIPSIDVFTFRC-----NCEGKT 1208
Db      1169 TT-----YVTTLAPASAPVTPATN-----AVPTTITT--TECSAATNAAGST 1207.

RESULT 7
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:9600117; PIDN:CAA84230.1; PID:9600118

Query Match          10.4%; Score 784; DB 2; Length 1188;
Best Local Similarity 29.9%; Pred. No. 3.4e-25;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

Qy      212 PTPPVPVVDAGSGLDNGDFKVTPTDSTTQHNKYSTPKITTAKPINRPSL---PPN 267
Db      460 PTPHSPPAD-----DVPPTPVPVPGKSPATSPSPQVPPASTPSPSLVKLSPPQ 510
Qy      268 SDTSKETSILTANKETVEKETTNNKQSTDGKEKTSKAKETOSIEKTSADLAPTSKV 327
Db      511 APVG-----SPPPVKTTSPAPIG---SPSPPPVSVV 541
Qy      328 -----LAKPTPKAETTTGKPLATTPKEPTPTTPEKAPATTTPKESAPTTTKEPA 382
Db      542 SPPEPVSPPPAVGVSPPEKSPPPAPVAVSPPPVKSPP--PPTLVAVSPPEPVKSP 599
Qy      383 PTTTKSAPTT--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 439
Db      600 PPAPVAVSPPPVKSPPPTPVASPPAPVAVSPPPVKSPPPPVSSPPEPKSPPPP 659
Qy      440 PKKAPATTTPKEPAPTT-----PKE-----PPTTTPKEPAPTTTKEPAPTTTKEPAPTTT 490
Db      660 PAKSTPPEEYTPPTSVKSSPPPEKSLPPTLLPSPPQEKPTPESTPSKP--DSEBEK 718
Qy      491 APTTPKEPAPTTPKAPATTTPKESPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 550
Db      719 SP--PKEVSSPPQTP-----KSSPPAPVSSPPTPVSSPPALAPVSSPEVKSSP----- 768
Qy      551 EPSPTTTPKEPAPTTPKKAPATTTPKKAAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKE 610
Db      769 PPAPLSSPPAPQVKSPPPPQVSSP--PPAPKSSPPAPVSSPPQVKTSSPPAPLSSPP 827
Qy      611 TTPKETAP-----TTP-----KULPTTPEKLAAPTTPKEPAP-----TTPPELAPTTPEEPT 657
Db      828 LAPKSSPPHVVVSSPPPVKSSPPAPVSSPEPLTPKASPPAHVSSPEEVKESPT--PA 885
Qy      658 PTPPEEAPATTPKAANTPKK--PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 716
Db      886 PTTYTISP--PSEPKSSPPTPVSLPPIVKSPPAMVSSP--PMTPKSSPPVVVVSSPPT 943
Qy      717 LKE--PA-----PTPKK--PAKELAPTTTKE-----PTSTSDKAPATTTPKGTAPTTP 762
Db      944 VKSSPPAPVAVSSPPATKSSPPAPVNLPPPEVSSSPPTPVSSPPAP--PKSSPPAP 1000
Qy      763 --KEPAPTTPKAPATTTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 821
Db      1001 MSSPPPEPVKSPPPAPVSSPPPVKSPPPAPVSSPP--PPVKSPPPPAPVSSPPPV 1057
Qy      822 --TPKETAPTTPKAPATTTPKKAAPTTPETPTTSEVSTPTTKEPTTIHKSPTDESPTPEL 880

```

Db 1058 KSPPPAPISPPPPVKSPPPPAPVS--SPPPPVKSPPPPAPVSPPPPPKSPPPPAPVS 1115
Qy 881 SAEPPTKALENSPKERGVPPTTKPATKEMTTAAKTERDRTTPET 930
Db 1116 SPPPPAP-----VKPPSLP--PPAPVSPPPPVTPAPPKKEBQSPPPPAS 1158

RESULT 8
T18535
high molecular mass nuclear antigen - chicken (fragment)
C1Species: Gallus gallus (chicken)
C1Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C1Accession: T18535
J.Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A1Title: A nuclear matrix-associated high molecular mass nuclear antigen, hNNA, of chick
A1Reference number: Z18955; MID:19803440; PMID:9365273
A1Accession: T18535
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: mRNA
A1Residues: 1-1151 <SH1>
A1Cross-references: EMBL:D88440; MID:d1177138; PID:d1025045; PIDN:BA24137.1

Query Match 9.2%; Score 695; DB 2; Length 1151;
Best Local Similarity 25.3%; Pred. No. 1.4e-21;
Matches 304; Conservative 108; Mismatches 466; Indels 324; Gaps 59;

Qy 105 AEVHNPTSPSSSK-----APPSCASQTIKSTKRSKPKPKKKTKKVISSEETEEH 158
Db 39 ARIH-PAAPQPPKWPVIGAPPPPTGTEPT-----PSKPTDGADAAPKASAEITSP 90
Qy 159 SVSENOSSSSSSSSSSSTIWKIKSKSAANRELQKLVKVNKKTKKKKPPKPV 218
Db 91 PASPPPPDPKAPSGAGAEAGTTPPSQGPAG-----TPP-- 126
Qy 219 VDEAGSGLDNGDPFKTTPPTSTTQHNKVS--PKITAKPINRPSLPNSDTSKETS 276
Db 127 -SGAAGAPKGDGTAPSGTSGADGKPAAGVPAKTTA--ATEAR-----ASAAP 176
Qy 277 TYNKETVETKETTTNK--QTSQDEKETSNAKESQTSKTSKADLAFTSKYLAKPTPK 334
Db 177 TVPKATAEATAVTAASQSAKATDAATAVTA--SQSAAPKATV--EVKPAATAVAKEAKA 232
Qy 335 AATTGKAPLTTPKEPTP--TTPKEPASTPKP-----TPTTKSAFTTKEPAPT-- 385
Db 233 VTAAAPAPATBAKAPATVSTIPICSSAAKELTAASPTASA--TAABAPVATASIM 290
Qy 386 -----TKSAPTTPEKAPATTTK-----BPAFTTPEKAPPT 415
Db 291 ATKVTAABAPAPSPVKATTTDKATATAPKAGPVKAVAVCAEAKAPPPPPQQLPK 350
Qy 416 TTKEPAPTTKSAPTT-----PKSPA-----PTPKKAPPTKEPAPTTPKEPTT 462
Db 351 AAAAAATCTELKPAAPAPHSFRANSHVTVTPPVPAATAATVP--TAGAVPKASTGT 408
Qy 463 T-----PKBAPTTKEPAPTTPKEPAPTAKKAPATTPKAPATTPEP--APTTKES 515
Db 409 TPAAPAPQGVPP--KAAPVTPSPQOAVPRAATAAA--APVTPQOVPYTKAATTNATP 461
Qy 516 PTTPEKAPATTKSAFTTKEPAP--TTTKSAFTTPEKESPTTKEP--APTTPKEP 568
Db 462 PQQPIPKATTTTATVTPQQRIPKAGTDAAPPAVPAKAPSGORATPPVPAATDPQRP 521
Qy 568 APTTPKAPPTPKAPAP-----TTPKEBAPT--TTKCAPATAPK-----EPAPT 611
Db 522 PTPPOSVPFAVTEPKQPPAAPPSPSEATPAVSPSPNLSPLPTPKFVPLMALVTPQGV 581
Qy 612 TPK-----ETAPTPPKULTPTTPEKLAFTTPEKAPATTPEEA-----PTTP-- 653
Db 582 TAAQMTQLAATKPSPLVPKASPK--ALMTTPPPPPGIPRLAALAAKLGIPSSVYASAMAK 640
Qy 654 --EEPTPTT-----EBPAPTTP-----KAAAPNTPE-----PAFTTPKEP 688

Db 641 VTPRLPASPVMMAASPASLGPDAARVALATNAASPGAKPBAAGNGTLMAPMAANTQM 700
Qy 689 APTTKEPAPTTPKETAPT-----TPKGTAPT----- 715
Db 701 APTGAAGAQTAPKMAATHVSPMGAGATQMSPTGANTHNSPIGAGATQMSPMGAAN 760
Qy 716 TLKEPAPTTPKAPKELAPTTKEP-----TSTSDKAPATTPKG--TAPTTPKEP 765
Db 761 TOMSPMGATTTQMSPMGAATTPQSPMGAAATQVNTSAGNTQMSPMGAATTPQTPSVG 820
Qy 766 APTTPEKAPATTPKGAPATTKEP--APTTPKAPKELAPTTKGP-----TSTSDKP- 818
Db 821 AATTP-QSPSPM-----GAATTLMSPMGAATTPQ--PSPMGAVTTPPMAATNTQPPPM 872
Qy 819 APTTKEP--APTTPKEP--APTTPKAP--APTTPPTP-----PTTSVSTPTTK 864
Db 873 AASTQSTPMGAATTTQSPMGATTTQSPMGASTPQAPPTVAGSPTPPPPIPSPTAQT 932
Qy 865 EPTTIHKSPESTPEL--SABPTPALENSPKERGVPPTTKPATKEMTTAKOKTER 922
Db 933 SPQMSKSPPPPPAPSAHAOTSPPAAVANAAPGV-TAVSPA--PIGVTAESPADGA 988
Qy 923 DLRTTPETTAPAKMT--KEATTTETTESKITATTTQVSTTTQDTPPKITTLKTTL 981
Db 989 RLSPPPTAATDGPASPAATADVEATD--VTAATAVPA-----EAP-----TK 1033
Qy 982 AKVTTTKKTTTITIMKPEETAKPKDRAINSKATTPPKQPTAPKPKPTSKKPKTMP 1041
Db 1034 AKRSSSSSSSSSSSSSSSSSSSSSSSDSSSSSSSESNPASPA-----P 1078
Qy 1042 RVRKPKTTPPKMTSTPELNPSTRIAEMQTTTRPQNTSNKLVENPKSEDAGAE 1101
Db 1079 AVGDQOQMTPGAASVPP-----VTEBAVQ-----EAAAAAAGAE 1117
Qy 1102 GE 1103
Db 1118 RE 1119

RESULT 9
T25697
hypothetical protein Fl6F9.2 - Caenorhabditis elegans
C1Species: Caenorhabditis elegans
C1Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C1Accession: T25697
R.Fulton, B.
submitted to the EMBL Data Library, August 1996
A1Description: The sequence of C. elegans cosmid Fl6F9.
A1Reference number: Z20071
A1Accession: T25697
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-1229 <FUT>
A1Cross-references: EMBL:U67956; PIDN:AA07691.1; GSPDB:GN00028; CESP:Fl6F9.2
A1Experimental source: strain Bristol N2; clone Fl6F9
A1Gene: CESP:Fl6F9.2
A1Map position: X
A1Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 9.0%; Score 676.5; DB 2; Length 1229;
Best Local Similarity 27.2%; Pred. No. 8.8e-21;
Matches 330; Conservative 95; Mismatches 385; Indels 405; Gaps 56;

Qy 73 RCFSEFGRBCDDCAQCKKYDKCCPDYESPCAEVHNPTSPSPSKAPPSGASQTIKST 132
Db 34 RDKFIVKHIKKQNTCTSC--KQVPD-----APSN-----PFDVSTTIS- 70
Qy 133 TKRSKPKPKKKTKKVISSEETEEHSVSENOSSSSSSSSSSSTIKIKSKKSAANR 192
Db 71 -----INNDVNDIGPSGDSNPTGSMQIEIATVGQGVSEHNIDSSV 114
Qy 193 ELQKLVKVNKKRKKKPTPKPPVVDAGSGLDNGDPFKTTPPTSTTQ--HNKVS 249

Db 115 EVEKVV-----TSDASITNAPTTGKSTT 140
Qy 250 PKITTA-KPINKRSLPPNSDT-----SKETSITNKEPTTETKETTNNKOTSDGK 301
Db 141 PEITIGIVINISKSBSVDMSTTRPSTLSPTELTSTPSTVSDSTST-EGTSPNPT 199
Qy 302 EKTSAKETSISAKDLAPTSVLAKPAPKATTTGKAPLTPKKEPTPTTREPAST 361
Db 200 TEISAPMETINTTETVTSVBSVSTLS--EDTITVTAINEST-----TTVAEVS 250
Qy 362 TPKEPTPTTIKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPA 421
Db 251 TTEBPTT-----ASTTKKSTT--KAPATTEPTPTTTEE--VTTTEAST 292
Qy 422 PTTKSAATTPKEPAATTP--KKAPATTPKEPAATTPKEP--PTTPKEPAATTPKEPA 476
Db 293 STTTSSESTSEK--PTPLIDNKIAGPATGK--PETHHPVIGTTPN--FDATETP 344
Qy 477 TTPKEPAATTPKEPAATTPKEPAATTPKEPAATTTKE-----PSPTTKEPAATTPKSA 531
Db 345 VAKSEDKMLSKTAAETTTQOTTEVT--DGEKETTANVSIPIPTVPLVETSTSTAS 403
Qy 532-----TTTKEPAAT--TTKSAPTTPKEPSPTTKE--PAFT-----TPK----- 566
Db 404 KESDGHITLKLKVTADSDSTESATTVKPNBEITTSHVPKTKGTVKVPKLELS 463
Qy 567--BPATTPKKAP-----TTTKE-----PAFTTP 589
Db 464 FDEPEIT-KAPHPKLEKTYHVLSDNAPRSEAKENDYMLDYNHREAKEPTT 522
Qy 590 KEPAV-----TTTKKAPTA--PKEPAPT--PKETAPTTKPLPTTPEKLA--PTTP 637
Db 523 EESSTBEVTTTBEANNGNPTTNPPTTTEOPTASTATTAALPTTEQVTTTBEPTTA 582
Qy 638 EKPAATTEELAPTTPEBPTTPEBPAATTPKAAAPTPKAPATTPKEPAATTPKEPA 697
Db 583 EKSTATO--KPTTGESVST--EKTSTTKA--STTEE--PTTDEPTT--ES 627
Qy 698 PTTKEKAPATTPKGAATTLKAPATTPKAPKELAPTTKEPTSTSDKAP--TT 753
Db 628 STTGAAATPELSTSEETTELKTTT--GSTTTEBPTTAPFAASTGIIYT 679
Qy 754 PKGAATTPKAPATTPKEPAATTPKGAATTLKAPATTP----- 794
Db 680 DEETSTSTSTPEITSTKE--VTESAITOTSVSVESSTPRLPERKKAIVNKEKMLE 737
Qy 795-----KKPAKELAPTTKGP-----TSTSDKAPATTPKETAATTPKEP 834
Db 738 VLKKEKRLKKEKSTSTGDSSETTVAENIDEVTTTEKVKVQTPTTKESTQEB 797
Qy 835 APPTPKKAPATTPPE--PTPTTSE--VSTPTTKEPTTHKSDPELSAEPPTKALEN 891
Db 798 TTTTTEKTSKTTTEKPTTSEATTEETTESPT--TEST----- 838
Qy 892 SPKRGVPTTPKAPATTPKEMTTTADKTEBDLRTPEPTTA----- 933
Db 839-----TVDTSKATTEESSTAETTTASB--TSETTSESAAFTGSPENTALQS 887
Qy 934-----APKMTKETATTTETKTESKITATTTQ--VTST--TTQDT 968
Db 888 SSQSENESSAEEKGARDPVPKHKHTVPAETTSVAASATTTTEITTTKSTTLET 947
Qy 969 TPKITTLTKTTTLAPKT--TTKTTITTEIMNKPETAPKORAINSKATTPKQ-- 1022
Db 948 TPIETATLINEVGAFTVGAPEDETITLIELSK-----INNTQISOQPTDIS 997
Qy 1023-----KPTKAPKPTSTKPKMPKRRKKTPTTPKMKSTMPLEIN----- 1063
Db 998 KTDALSSISGLISFTKAPAPTI-----HTTDAAPVATATASLNDGSDK 1044
Qy 1064-----PTSRIAEA 1071

Db 1045 KIIDEAQPTDEIRRA 1059

RESULT 10

A:5175

muscin 1 precursor, repetitive splice form A [validated] - human
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistatin,
nucleic acid; mucin; polynuclear epithelial mucin (PM)

M:Contains: muscin 1 precursor, epithelial tumor antigen splice form; muscin 1 precursor, e

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000

C:Accession: A:5175; B:5175; A:5886; A:5887; S:10572; S:40293; A:6735; E:00066; S:10218; S:10

R:Liteneberg, M.J.L.; Vos, H.L.; Gemmessen, A.M.C.; Hilkmans, J.

J. Biol. Chem. 265, 5573-5578, 1990

A:Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene encc

A:Reference number: A:5175; MUID:90202794; PMID:2318825

A:Accession: A:5175

A:Molecule type: mRNA

A:Residues: 1-952, 1033-1344 <LI61>

A:Cross-references: GB:M32739; GB:005288; NID:g182121; PIDN:AAA5804.1; PID:g182124; GB:M

A:Experimental source: Eplife form A

A>Note: Genbank entries HUMEPIS1A1 and HUMEPIS1A2 present only the amino- and carboxyl-enc

A:Accession: B:5175

A:Molecule type: mRNA

A:Residues: 1-19, 29-952, 1033-1344 <LI62>

A:Cross-references: GB:M32739; GB:005288; NID:g182126; PIDN:AAA5806.1; PID:g182129; GB:M

A:Experimental source: Eplife form B

A>Note: Genbank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino- and carboxyl-enc

R:Gendler, S.D.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchell

J. Biol. Chem. 265, 15286-15293, 1990

A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithelial

A:Reference number: A:5886; MUID:90368715; PMID:1697589

A:Accession: A:5886

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19, 29-992, 1033-1344 <GEN>

A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59676.1; PID:g188870

A>Note: Genbank entry HUMUPAMU includes one copy of the tandemly repeated sequence

R:Plan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.

J. Biol. Chem. 265, 15294-15299, 1990

A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A:5887; MUID:90368716; PMID:2394722

A:Accession: A:5887

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-19, 29-1109, 'S', 1111-1339, 'A', 1341-1344 <LAN>

A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599

A>Note: Genbank entry HUMUPAMU contains four fewer copies of the tandemly repeated sequen

R:Wreschner, D.H.; Hareven, M.; Tsarfacy, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.

Eur. J. Biochem. 189, 463-473, 1990

A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may generat

A:Reference number: S:10571; MUID:90276413; PMID:2351132

A:Accession: S:10572

A:Molecule type: mRNA

A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WRB>

A:Cross-references: EMBL:X52229; NID:g37053

A:Reference number: S:40293

A:Submitted to the EMBL Data Library, March 1990

A:Accession: S:40293

A:Molecule type: mRNA

A:Residues: 1-19, 29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <WR2>

A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA6478.1; PID:g37054

R:Abbe, M.; Siddiqui, J.; Kule, D.

Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated

A:Reference number: A:6735; MUID:90088473; PMID:2597151

A:Accession: A:6735

A:Molecule type: mRNA

A:Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>

A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA5757.1; PID:g181543

R:Maenawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H

J. Biochem. 112, 609-615, 1992

A>Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
A/Accession: JX0235; MUID:93123189; PMID:1478919
A/Accession: PX0066
A/Molecule type: mRNA
A/Residues: 998-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
A/Experimental source: gastric carcinoma cell
A/Strain: Licht, S.; Baruch, A.; Broy-Stein, O.; Kedar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A>Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
A/Reference number: 851026; MUID:95080414; PMID:7988707
A/Content: annotation
A/Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
C/Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
partial repeats. The repeat shown is defined by SmaI nuclease sites.
C/Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
C/Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C/Genetics:
A/Gene: GDB:MUC1; PUM
A/Cross-references: GDB:120705; OMIM:158340
A/Map position: 1q21-q23
A/Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C/Superfamily: polymorphic epithelial mucin
C/Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphic
P/1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRA>
P/1-62/Region: mucin 1 amino-terminal non-repetitive
P/1-23/Domin: signal sequence #link PRA #status predicted <SIGA>
P/1-19, 29-32/Domin: signal sequence #link PRA #status predicted <SIGB>
P/1-19, 29-32/Product: mucin 1 precursor, splice form B #status predicted <PREB>
P/1-19, 29-32, 1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
P/138-1017/Region: 20-residue repeats (GSTAPAHGVSADTRAP)
P/1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
P/1245-1272/Domin: transmembrane #status predicted <TRM>
P/1046, 1064, 1118, 1144, 1222/Binding site: carbohydrate (asn) (covalent) #status predicted
P/1213/Binding site: phosphate (tyr) (covalent) #status predicted

Query Match 8.9%; Score 672; DB 1; Length 1344;
Best Local Similarity 27.8%; Pred. No. 1.5e-20;
Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

QY 276 LVNKKETVETKETTNNK-QSTNGKEKTSAKTOSIEKTSKD-LAPSKVLAKPT 332
DB 15 LVVLATTAAPKAVTVTGSHASTPGCEKETSATORSVPSSTENAVSMTSVLSHS 74

QY 333 P-KAETTTGP--ALTTKEP-----TPTPKESASTP-----KE 365
DB 75 PGSGSTTGQDVTTLAPATEBPASGSAATGQDVTSVTRPALGSTTPPAHDVTSAPDNK 134

QY 366 PPTTT-----IKSAPTTKEPAPTT-----KSAPTTKEPAPTTKEP-----A 405
DB 135 PABGSTAPPAHGVTSADT--RAPGSTAPPAHGVTSADT--RAPGSTAPPAHGVTS 190

QY 406 PTPPKPAPTTTKEPAPTTTKSAPTTKEPAPTT-----KKAPTTKEP----- 451
DB 191 PPT--RAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGV 247

QY 452 --APTTKEPTPTT-----KEPAPTTKEP-----APTTKEPAPTAPK----- 488
DB 248 TGAAPTDRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGV 307

QY 489 KPAPTTKEPAPTTKEP-----EPAPTTKEPSPTTKEPAPTT-----TKSAPTTTKEPAPTT 541
DB 308 TSAAPTDRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGV 367

QY 542 TKSAPTTKEPSPTTKEP-----APTTKEPAPTTKEP-----KPAPTTKEPAPTTKEP 590
DB 368 T-SAPPT--RAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPA 424

QY 591 ---BPAPTTTKKAPAPAPKAPAPATTEKETAPTTKLGTPTTPEKLAPTTPEKAPAPTTP- 646
DB 425 HGVTASAPDRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPA 484

QY 647 ---ELAPTTPEEPTPTT-----EAPAP--TTPKA---AAPNTPKESAPPTPK- 686

DB 485 HGVTASAPDRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPA 544
QY 687 ---BPAPTTTKEPAPTTKEKETAPTTKGTAPPTTLKEPAPTT-----KKAPAKEL 733
DB 545 HGVTASAPDRAPAGSTAP-----PAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGST 600

QY 724 APTTKEPTTSTSDPAP--TTPKGTAPTTKEPAPTTKEPAPTTPGK-----TAPTTLK 787
DB 601 AP-PAHGVTSADTTRAPAGSTAPPAHGVTS-----APDRAPAGSTAPPAHGVTSADTTR 655

QY 788 BPAPTT-----KKAPAKELAPTTTGTSTSDKAPAP--TTPK-----ETAPTTT 831
DB 656 APGSTAPPAHGVTSADTTRAPAGSTAP--PAHGVTSADTTRAPAGSTAPPAHGVTSADTTR 714

QY 832 KEPAPTT-----KKAPPTTPEPTTPTSVSTPTTKEPTT-----IHKSPD 874
DB 715 PABGSTAPPAHGVTSADTTRAPAGS--TAPPAHGVTSADTTRAPAGSTAPPAHGVTSAD 772

QY 875 ESTPELSAPTPPKALENPKERGVTPTTPAA-----TKPEWTTAKOK--TTERD 923
DB 773 TRPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSAD 832

QY 924 LATTPETT-----TAAPKMTKETATTTKTESKITATTTQVSTTODTTP--PKIT 974
DB 833 TRPAGSTAPPAHGVTSAP-----DTRPAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGV 888

QY 975 TLKTTTLAPKTTTKTTT--EIMKPEELAKKDAATNSKATTPKOKTAPKAPT 1032
DB 889 SAPDRAPAGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAP--GSTAPPAH 946

QY 1033 STKKKTPVRKRPPTTPTPRMTSTPELNP-----TSRIEAMLTQ--TTRP--NOTPN 1084
DB 947 VTSADTTP--APGSTAPPAHGVTSADTTRAPAGSTAPPAHGVTSADTTRAPAGSTAP 1003

QY 1085 SKLVEVNPKSDAGAGETPHMLRPVFMPEVTPDMDYLPVNOGIIINPMUSDEN 1144
DB 1004 AHGVTSADTTRAPAGSTAPPAH-----GVTSADPNRPAAGSTA--PVHANT 1049

QY 1145 IONGRPVUGLTLRNGTLVAFRGHFMMLSPSPS 1180
DB 1050 ASGSASGASSTLVHNGTSARATTTTAPSKSTPFSIPS 1085

RESULT 11
138346
elastic titin - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C/Accession: 138346
R/Labell: S.; Kolmerer, B.
Science 270, 293-296, 1995
A>Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A/Reference number: A57430; MUID:96026330; PMID:7569978
A/Accession: 138346
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-7962 <RES>
A/Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427
C/Genetics:
A/Gene: GDB:TTN
A/Cross-references: GDB:127867; OMIM:188840
A/Map position: 2q31-q31

Query Match 8.9%; Score 666.5; DB 2; Length 7962;
Best Local Similarity 24.2%; Pred. No. 1.4e-19;
Matches 268; Conservative 120; Mismatches 491; Indels 229; Gaps 50;

QY 98 PDYEFCAEVNPTSPSSKAPPPSGASQTIKSTKS-----PKPPKK 143
DB 6827 PEEEFVPE--EVLPEVKVPVPAVPELIKKTTEKVVIPKKEAPPAKVPEVPKV 6884

QY 144 KTKKVI--ESEEITEHVSSENOESSSSSSSSSTIWKIKSKNSKANRELQKCLKV 201


```

Db 6885 EKKRIILKEBEVLVEVEEPEEPISE-----EPEEPPSIEVEEVAAPRV 6935
Qy 202 DNKKRRTKKKPPKPPVDEAGSGLDN--GDKYKTPP-DTSTQNKSTSPKITTAP- 257
Db 6936 EVIKKAVDEAPTPVKKVAPPAKAKVSKKIPEEKVAVPQKKKAPPAKVEVKVPEKVV 6995
Qy 258 -INPPSLPP---NSDTSKETSLTVNKKETVETK---EYTTTNKQTSIDGKE----- 302
Db 6996 LVPKKEAVPPAKGRVLEBKSVARQEVVKKERLELEVAEVAEVEIEEPEEFHVEEYF 7055
Qy 303 -----KTSAKETQSIKT-----SAKDAPTSKVLAKPPPKAETTKG 341
Db 7056 EGEHEVEEVEEPIKLEQHVEEHEHRAKVRVLEVEAESEVEFEKPPAP-----KG 7107
Qy 342 PAUT-----TPKEPPTTP-KEBASTTPKEPPTTIKAPPTPKP-APTTKSAPTTPK 394
Db 7108 PEISEKIPPKKPPPKVVRKPEKPAKVPBVPKKIVVEKVRVVEEPRVPTKVPEVLPPK 7167
Qy 395 EPAPTTKEBAPTTTPKEBAPTTTPKESAPTTTPKESAPTTTPK-----PAPTPE 450
Db 7168 EVVP-----EKKVVPAPKKEAP-----PKVVEAPKEVVEPKVPVPPKKEV 7213
Qy 451 PAPTKEPPTTPPTTPKEBAPTTTP-----KEBAPTPAKKAPPTTPKEBAPTT 503
Db 7214 PPTKVPEVPKAVPEKKVPEAIPPKESPPPEVEEPEESPPAPK-KPEVPVAVPEVP 7272
Qy 504 KEBAPTTKEBPTTPKEBAPTTTPKESAPTTTPKESAPTTTPKESPTTPKEBAPT 563
Db 7273 KEVVEKKVAPAP--PKKPEVTPVK-VPEAPKEVVEPKVPV--PKKPEVPTK----- 7323
Qy 564 TPKEBAPTTTPKAPTTTPKEBAPTTTPKAPTTTPKAPTTAPKEBAPTTTPKETA 623
Db 7324 VPEVVKAVPEKKV-----EAIPPKESPPPEVEEPEEVALIEEPPAVVEEPPAPQ 7379
Qy 624 LT-----PTTPEKAPTT-----PEKAPTTPEELAPTTPEEPTP-TTPEEAPTTPKAAP 674
Db 7380 VTPVPEKAVPEKKAVVAKKPELPPVVKVPEVPEKVPVAVPK--PEAPKAPV 7437
Qy 675 NTPKEBAP-----TPKEP-----APTTPKEBAPTTTPKETAAPTTTPKEBAPT 724
Db 7438 EVPEKVEVPEKKAVVAKKPEVPAKVPVPEKVP-----LEEKAVP--VPEAPESPPREV 7490
Qy 725 PKKAPKELAPTTTPKEPTSTSDKAPTTTPKGTAPTTTPKEBAPTTTPKGTAPT 784
Db 7491 YEB--PEEIAP--EEBIAPEEKVPVAE-----EBEVPPEAPPEBKIIPEKVP- 7540
Qy 785 TLKEBAPTTPKKAPKELAPTTTPKGTSTSDKAPTTTPKETA-----PTTPKEBAPT 837
Db 7541 VIKKKEAPPPKPEPEKVP-----IEKPKLRPPPPPPAPPEKEDVKEKIFOLKAI 7597
Qy 838 TPKEP--APTPEPPTTPPTSVE--TPTTKEP-----TTIHSPPESPPELSAPTP 886
Db 7598 NPQVEKVELTPKVPGEKKVRKLLPERKEPKEEVLKSVLRPEEPEKVE---P 7653
Qy 887 KALENSPKPEGVPTTPKTAATKP-----EMTTAKDTERDLRTPETTTAPKTKETA 942
Db 7654 KKLK-KVKKAPV--EBPPKPVVEVAVPTTKERKIPETKVPLEKPAIPLAPBPX 7709
Qy 943 TTTKETTESKITATTTQVSTTTQDTPPKITLKTLLAPKVTTKTITTTIEMK- 1000
Db 7710 PKPEA-----EVTIKPPEVEPEPTPIAIPVVPVVGKAAE 7745
Qy 1001 -----PEETAKPKDARTNSKATTPKQKPTKAPKKTTS--TKPKTMPVRKPKTTPP- 1052
Db 7746 AKAPKEEAKPPGPIKGVPKKTPSPIEARRRLRGSGGEGKPPDEAPPTTQKAVPLKV 7805
Qy 1053 ---RKMTSTMPELNPTSRIAEAMLOTTT 1077
Db 7806 KEIKDIIITSEBFVSSAIFECIVSPST 7833

```

RESULT 12
745025

```

mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: U45025
R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Iaine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A>Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: 228999; MUID:97166151; PMID:9013550
A:Accession: U45025
A>Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:Z72496; NID:G1834502; PIDN:CAA96577.1; PID:G1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match. 8.8%; Score 663; DB 2; Length 3570;
Best Local Similarity 22.1%; Pred. No. 8.7e-20;
Matches 396; Conservative 119; Mismatches 578; Indels 698; Gaps 72;

Qy 54 MECCPDPRVCTAALSCKGRCFESFERGECDCDQCKDCCPYE--SFCAEV-HNP 110
Db 1581 VECSLDGLVC-----RNRE-----QVGR-KMCFNYEIRVFCNNGHCP 1619
Qy 111 TSPSSKAPPS-----GASQIKSTTKRSPKPKKTKKYIE 150
Db 1620 STPATSTPATSSPTGTTWILTEQTATTAATTAGSTALPSSTGTAPP-----KULT 1673
Qy 151 SEETHEESHVENOESS-----SSSSSSSSSTIKWS-----KNSANR--- 192
Db 1674 SQATTPTATSKATSSSPRATTLPLVLTSTATSTATSPFIPISSTLGTGTSQNNP 1733
Qy 193 -----ELQKLVKONKKNRTKKKTPPKP-----VDE- 224
Db 1734 PMATMSTIHPSSTPEHTHTSTVLTTKATTTTATSTMSSTSPGTWILTELTAAATTA 1793
Qy 225 GLDNG-----DEKVTTPDSTTQNH-----KVSTPKITTKAPI 258
Db 1794 ALPHGTSPSTGTGTTWILTEBPTTATVTPGCTATSTASTATAGTAKVLIS-TATTPVI 1852
Qy 259 NFR-----PSLPNSDTSKETSLTVNKKETVET-----KETTTNKQTSIDG 300
Db 1853 SSRATPSSPGTATALPALNSTATTPTATSTVTAIPSSLSIGTAWRLSQTTPTATMSTAT 1912
Qy 301 KEKTSKETSIKTSKADLAPTSKVLAKTPPKAETTK-----GPA 343
Db 1913 PSTPTVHTSTVLTATTAATTTTGTGSAVAPSSSTPGTAHTTKVPTTTTGFTATPSSSPGTA 1972
Qy 344 LTPP-----KEPTPT-----TPKE-----PASTTPKEPTP 368
Db 1973 LTPPVWISTTTTPTRGSGTVPSSIPTGTHATVLTITTTTAVATGSMATPSSSTQTSPT 2032
Qy 369 TTIKAPTTKEBAPTTTPKAP--TPKEBAPTTT-KEBAPT-----TPKEBAPTTKEPA 421
Db 2093 PSLTTTATTATTAAGSTINPSSTPGTPIPLVLTATTPAATSTVPSALGTHHPV 2092
Qy 422 PTTKSA-----PTTP-----KEBAPTTPKKAPPTT--KEP 451
Db 2093 ENTATTAHGRSLPPSSSHVPTAATSAISGLGTHTEBETGSHTPAATTTGTTQSTP 2152
Qy 452 APTTPKEPTPTTPKEBAPTTTPKEBAPT--TPKEBAPTPAKKAPTTTPKEBAPT--PKEBAP 508
Db 2153 ALSSPHSSRTTESPSPGTTTPTHTGTGSKTTTATATPSKTRISTLLPSSPTSAPIITTV 2212
Qy 509 TTKEPS----- 515
Db 2213 TTGGEPCAMSEWLDYVPMGPBGDPDTYSNIRAAGAVCEQPLGBCAQAQPGVPL 2272
Qy 516 -----PTTPKEBAPTTTPKAP 531
Db 2273 RELQGVVECSLDGLVGRNBEQVGKFMCFMCFNYEIRVFCNNGHCPSTPATSTPATSTP 2332

```


Qy	532	TTT-----KEBAPTTKCAPPTTKEKPSSTTTKEBAPTTKPKCAPPTTKEKAPPTTKEBAPPTT	583
Db	2333	GTTHLTLTKLTTATTTATTSIGSTATSSITQGPAGTPh--VSTIATPTTGVTSK-----ATP	2386
Qy	590	KEBAPTTTKKPA-----PTAPKEPA-----PTPKETAATTPKCLP-	626
Db	2387	FSSPETATALPALRESTATPTATPSSTLIGTWRRLSQTTPhMATMSTATPSSSTPE	2446
Qy	627	-----TTPEKLAPTTPEKAPTT-----PEELAPTTPEEP--	656
Db	2447	TVHTSTVLTTATTTGATGSAVATPSSSTPGTAHTTKVPTTTTGTFTVTPSSSPGTAHTTPV	2506
Qy	657	---FTTPEEBAPPTPKAAAPVTPEKAP-TPKEBAPT---TPKEBAPT--TPKE--T	704
Db	2507	WISTTTTTTSGSIVTPSSISPEOTHTPTVLTTTTPAVATGSMATPSSSTQTSIGTPRLIT	2566
Qy	705	APTTPKGTAPTTLKEBAPTTPKKAPKELAPT-TPKEPTS-----TTSOKPAPT	753
Db	2567	TATITATGSGT--NPSSTPGTPIPELITTTATTPAATSSVTPSSALGTTHTPVBNT	2622
Qy	754	PKGT-----APTTP-----KEBAPTTKEBAPTTPKGTAP-TLK	787
Db	2625	TATHTGRSLSPSSPTTAVTATSATSGTLGTHITPESTGTSHTPAAITGTTTTSTPALS	2684
Qy	788	EPAP---TTPKKAPAKELAPTTTKGPTSSSDKAPPTTPEKAPTT--PKEP-----	834
Db	2695	SPHPSRSRTTEBPSP---GTTTPEHTATISKTITATPSPKTRISTLLPSQTSAPITVV	2740
Qy	835	-----	834
Db	2741	VTTGCEPQCAWSEMLDGYPMPPGSGDFDTYSNIRAAAGAVCEQPLGLECRATAQGPV	2800
Qy	835	-----APTTPKAPPTTEPT	850
Db	2801	LGELGQVECSLDfGLVGRNBQYQKFKMCFNYEIRVCCNYGHCPSPTATSSSTAMPST	2860
Qy	851	PPTT---SEVSTPTTKEPTTIHKSPDESTEPBLAPPKALENSPKKPGVPTK-----	902
Db	2861	PCTTWILTELTTATTTATTSSTGTAAP--SSTP--GTAAPPKVLSTBATTTATTSKATSS	2917
Qy	903	-----TPAATK-----DEMTTAAKDTBERDLRTPETT	931
Db	2918	SPRTATTLPVLTSTATKSTATSVMPIPSSTLGTTGLBEQTTTPAATSTIHPSTPEPT	2977
Qy	932	-TAAPKMKETATTEKT-----TESKATATTTQVPS--TTQOTTP-----	970
Db	2978	HTSTVLTKATTRAITSSTSPSSPTGTWILTELTTAATTTAGGPTATPSSSTPGTWIL	3037
Qy	971	FKITTLTKTTLAPKVTYT--KTIITTEIEMNPEE---TAKPKORAT--NSKATPKPQ-	1022
Db	3038	TELTITATTTASSTGSTATLSSPTGTTWILTEBPTATTAATPAGSTATSSQIATGPHV	3097
Qy	1023	-----KPTKAPKKTSTYKPKT--MPRVKPKTTPPKMT-----	1056
Db	3098	STIATTPVTSSKATPSSSPGATLALPALRSTATTPTATSPALPSSSLGTTWRLSQTT	3157
Qy	1057	-----STMELNPTSRIAEMLQTT-----TRNQTTPSKLVEVNPKSBDG-	1098
Db	3158	TPATMSTATPSSSTPEVHTSTVLTTATTTGATGSAVATPSSSTPGTAHTTKVPTTTTGF	3217
Qy	1099	-GABGETPHMLRPHVMEPTVPMDDYAPRVNOGIILNPMLSDETNIENCKPVDGLTTL	1157
Db	3218	TATPSSSGCATLTPPVWISTTTTPTTTTP--TTSSIVTPSSISPEOTHTTAVALTTTTTY	3275
Qy	1158	RNGTLVAFRGHYFMWLSPPSPPARRITEVWGLPSPIDIVPTRCNCEGKT	1208
Db	3276	ATGSM-----ATPSSS---TQTSIGTPSLTTATTTATTTAGST	3309

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: J14513
R:Ravetto, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z1536
A:Accession: J14513
A:Stratus: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GNO0021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/2; 1175/3; 1207/1; 1409/2;
3504/1

Query Match	8.7%	Score 658.5	DB 2	Length 3507
Match Local Similarity	22.6%	Pred. No. 1.3e-19		
Matches	286	Conservative 178	Mismatch 484	Indels 317; Gaps 51;
Qy	27	LSSCAGRCGEGYSRD-ATC-----NCDVNCOHYMECCPEFKVCVTAELSCRCRESF	78	
Db	1781	MGSGCGCKCMAGYGDATCTCKIEEPRSDXTA-----CTDEMRCLGLE---KKQCTVDE	1832	
Qy	79	ERGREC-----DCAQCKKDKCCPDYSSFCAEVHNPTSPSSKAP	121	
Db	1833	EEVPOCGACLPGHHPINGTCOSIQISGLCAQKDCNHNAE--CIDIH-----PDSHFCSC	1885	
Qy	122	PSG-----ASQITKSTKSPKPNKKTKKYIESEIT	155	
Db	1886	PDGFTGGMCDVDVDECNAGMCDDENTKCENTIGSNVCYCLSEFKVYDEKCVDEKKOP	1945	
Qy	156	EEHSVSENQSSSSSSSSSTIWKIKSSKNANNELOKUKVKDKNKRTKCKPTK	215	
Db	1946	NREKIEIDEENSSSNNG-----QEKPTTK	1970	
Qy	216	PPVYDEAGSLDNGDFKVTTPDITSTTQHNKVSFPKTTAKPINPRSLPN--SPTSK	272	
Db	1971	GIYVST-----ATSESTTAEPHYTISSTTKMTSKSEPNVTMSSESP	2019	
Qy	273	ETSLTVKERTV-ETKEETITNKQTSIDGKEKTTSAETQSIIEKTSAKDLA---TSKV	327	
Db	2020	EVSTSSSKSTTASSETIYVSTPSSSSSEALTSPPATTTIVITSSYKSTTPKKESSSEI	2079	
Qy	328	LAKETPAEFTTNGPALTTPEKEPTPTTPKEBPASTTPEKPTTIKS-APTTPKEPAPT	386	
Db	2080	TVKLSSKSPREVTSSVSSSPSTPS-TTSQVTSIVPETSXYLSSLPDVSSTTPENIT	2138	
Qy	387	KSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTASAPTTTKEPAPTTPKKAPT	446	
Db	2139	-SSSTKSLSSASTGDTNSTTSTSSLASVKSTABEGTSASVAPAKLSLSPDVQSFS	2197	
Qy	447	TPK-EPAPTTPEKEPTPTTPKEPAPTTKEP-APTTPKEPAPTAAPKAPPTTPEKAPTPK	504	
Db	2198	TKTFDAESTTVQASHTSSSTSYKSTSEPSNHTKLSTISNSNSSSPVTSFKSTPVPPE	2257	
Qy	505	EPAPTTTKEPSTTP-KEPAPTTTASAPTTTKEPAPTTKAP-----TTPKPSPT	556	
Db	2258	-----STEQPTSTPSSQSLTPMNSNSEVLTTEPHLSSLSLPDVQSSTTPNNLSSESS	2312	
Qy	557	TKEPAPTTKEPAPTTPKKAPPTTKEPAPTTTKEPAPT-----	596	
Db	2313	TVE-----TPTTSSSVSLNSEPSITTEAPTLSPDILSTTNNLSQSSTVSTEDRSIISSE	2368	
Qy	597	-TKKPAPTAK-----EPAPTTPEKETAETTPPKLPTTPPEKLAAPTPEKAPAPT	645	
Db	2369	NSEKPT-SAEELVTSVTHVASSDPVPTSS--SEPDGLGSSGTENIPKASSQOTISSTP	2425	
Qy	646	BEAAPTTPEEPTPTPEEP-----APTTPKAAANTPEKAPTT-----	684	
Db	2426	TPDITTAASEPFTKSTMSPLSTSTSNVLSSESTTPBESS---KSPVSSSTEGISVTVST	2481	

QY 685 --PKRAPL-----TPKEAP-----TPKEAPTPKGTAPTLKEAPTPPKP 728
DB 2482 EFSKPEESTISSVLEEDLTKTPSPLEETTPASSETSEPLEDSLTVRIHELTTSSN 2541
QY 729 APKELAPPTTKKPTSTSDKPA-----PT-----TP--KGAP- 759
DB 2542 VPKESESTTSSESSKPSGEPAGILTSVTVPTSSVSLTASEIAITSNPFKGRPI 2601
QY 760 -TTPE-----EPAPTPKEAPTPPKGTAPTLKEAPTPPKAPKELAPTTKGP 811
DB 2602 TTPSKSLVKSSTTSPTVTSSEPSSESTKRTVSTVSTTPTEETSSSLITTAAPSKPT 2661
QY 812 -STSDKAPPTPKKAPTPPKAPTPPKAPTPPETTPPTSEVSPPTTKAPTIN 870
DB 2662 ESTTSSAPPTTPAKTSETKPSNVSTSKSTENVTSTSGSLSSSTMS- 2715
QY 871 KSPDESTP--ELSAEPTPALE--NSPKBPVPTTKPAATK-----PEMTTAKDKTTE 921
DB 2716 SEPETNAPAVTVSSASSTLEENSTSSP-----TSSEASVLSLFPESITSEAVTVSS 2771
QY 922 RDLKTPETTTAPKMTETATTKTESKITATTT-----QVTSSTTQ 966
DB 2772 R-----APAEITWSSSHREISTVSESPSEPEIPSTVSPNVVTASSIPSEEPILSYTS 2827
QY 967 DTPP-FKITT-----LKTTLAPKVTTTKITTEIMNKPEETAKPDRATNSKATPK 1020
DB 2828 STTPVRLITGTPDILVSVTVSHGNRRONITASSV--PSNSTPILPSES--LTPQ 2883
QY 1021 POKPTAKPKPTSTKPKTPRVRKPK--TTPTRKMTSTWPELN-----PTSRIBAM 1072
DB 2884 PPTTTTAKPATSTGKKGPPSIOCPAEMFTTPAP-----PPSNGVGSEETNGEBOV 2937
QY 1073 LQTTT 1077
DB 2938 TSTTT 2942

RESULT 14

151618
nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: I51618; #55757
R:Calms, C.; McStay, B.
J:Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: I51618; MUID:96019267; PMID:7593294
A:Accession: I51618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C:Gene: xNopp180
C:Superfamily: nucleolar-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 8.6%; Score 643.5; DB 2; Length 990;
Best Local Similarity 24.2%; Pred. No. 1.6e-19;
Matches 246; Conservative 148; Mismatches 424; Indels 159; Gaps 40;
QY 112 SPPSKKAPPPSG-----ASQIKSTTKSRKPKPKKT-----KYIESREITEE 157
DB 57 SPDARKKRPANGLPKKSAKESSESSSEEDPPAKKAQPAAGKPPVKAQVPPKA 116
QY 158 HVSVENOSSSSSSSSSSSTIWK--IKSKNSAARELOKCLKVNDKNKNT-----KK 210
DB 117 KSSSEDSDESDSEETTKKPAKRAQTPKAAVVTTPQKAKASSSSSSSEDSASAKK 176
QY 211 KPTPKP--VVDAGSGLDGDFVTTPTDSTGHNKVSSTPKITTKAPINPRSLPN 267
DB 177 QPVIKVPKQAVVAKGLASNG--KTADSSSSSDSDSPAKTKAATK-----TPP- 224

QY 268 SDTSKETSITVKNKETVETKETTNNKQTSNGKKTTSAKETOSIEKTSAKDLAPTS-- 325
DB 225 --TPRAITAKQAKTKAGKSSSSRESDSDSDEQKTKSKPKPDVSAVP--PPTSVS 279
QY 326 --KVLAKEPTPAETTTKGPALTTTPKEPTTPPKK--PASTTPKEPTPTTIKSAPTPKP 381
DB 280 KKTLLSQGTAKAESSDSDSDEBQPAKKAIVPAKKAASAPKLAKAETSTOSES 339
QY 382 APPT-----TSAPTPKEAPTTKEAPTPPKAPATTKE--PAPTTISAP 429
DB 340 DSSSEDEKSSVKLGVKAAP--KKAFA-----APAKSTPVAAKASAPAKASSSSD 390
QY 430 TTPKEAPTPPKDAPTPKEAPTPPKAPTTTPPKAPTPPKAPTPPKAPTPPKAP 488
DB 391 SDSSNEETTTKPAKATTPAKSA--ATPSTKPTNGK--ATPSTKPAKAGTITKATKDS 448
QY 489 KPAETTPKEAPTPPKAPTPPKAPTPPKAPTPPKAPTPPKAPTPPKAPTPPKAP 548
DB 449 SSSSDSDSDEETTTKPAATTPAKSAATPTSKPTNSKATPTSKTPAKGTPKTSAA 508
QY 549 PKESPTTKAPATTPKEAPTPPKAPTPPKAP--APTPKEAPATTTKAPAPAKE 607
DB 509 KKDSSSDSDSDSDEKTPA--KRAAKTPPAKPAKTPAKPA-----AKTPAK 557
QY 608 PAPTPKETAFTTKKLLPTTPKLAFTPEKAPAPTPPELAPTPPEPTTPPEAPPT 667
DB 558 PA-----AKSTPKQVPTKE-----SSSDSSSSEDEKKS 590
QY 668 TPKAAPTPKEAPTPPKAPTPPKAPTPPKAPTPPKAPTPPKAPTPPKAPTPPK 727
DB 591 SAKAVKTT--PKATSK--PVVASKPV--AKKASSSDSDSEETTKTTPPLTGLS 643
QY 728 PAPKELAPTTKEPTSTSD-----KPAETTKGAPTPPKAPAPTPPKAPTPPK 780
DB 644 PAVXTLPKKAESSSDSDSDESKTPAPKAPKASATPVNTKAPAONKASKASCSDS 703
QY 781 TAPTLKEAPTPPKAPKAPKAPLAPTTKGPSTSDPAPTPKETAFTTPKEAPTPPK 840
DB 704 SSEEKSKOP--TKSPAKATAP--PKNPVAVNKDPSSSSSDSDGDDKQ--KPK 757
QY 841 KPAETTPETPTTSEVSTPTTKEPTTIHKSPPDESIPBELSAEPTPALENSPKRGPVT 900
DB 758 QAA-----AAKOVQOAKAKKPTPKKAASSSE--DS 787
QY 901 TKTPAATPEMTTAKOKTERDILTPETTTAPKATKETAATTKTESKITATTTQV 960
DB 788 SSDEVSKAKKTNTAVSKSPV--TTPKAVPAKKSSESSSEDEKQGNKNTSTTKI 843
QY 961 TSTTQDTPPKITLAKTTTLAPKVTTK--KITTEIMNKPEETA--KPKDRATNS-- 1014
DB 844 ANST-----PKAAAACGSESSSEDEKANGTSGKRGESTGNAEC 885
QY 1015 KATTPKQKTPAKKPTSTKPKTPRVRKPTTTP--RKMTSTWPELNFTSRIAE 1070
DB 886 EAVTPE-----NKDLKAKSPNTFPKVNKELKNTPPRVVDEIDIEINP--RMAD 932

RESULT 15

711622
extensin class 1 precursor - cowpea
C:Species: Vigna unguiculata (cowpea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T11622; #54155
R:Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
Mol. Plant Microbe Interact. 10, 95-101, 1997
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.
A:Reference number: Z17301; MUID:97155574; PMID:9002273
A:Accession: T11622
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-489 <ARS>
A:Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2003, 09:20:36 ; Search time 18 seconds

(without alignment)
3668.083 Million cell updates/sec

Title: US-09-556-246-1

Perfect score: 7526

Sequence: 1 MAMKTLPIYLTLISVFIYQ.....ARAITTRSGQTLISKYWNCP 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1188	15.8	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	12.6	1664	1 SLPI_CLOTM	Q06852 clostridium
3	809	10.7	1367	1 AMYH_YEAST	P08640 saccharomyc
4	707.5	9.4	555	1 GPI_CHLRE	Q9F966 chlamydomon
5	703.5	9.3	5085	1 PCLO_RAT	Q9JKE6 rattus norv
6	665	8.8	5038	1 PCLO_MOUSE	Q9GTX7 mus musculu
7	660	8.8	5703	1 MUSB_HUMAN	Q9HCB4 homo sapien
8	659.5	8.8	5147	1 PCLO_HUMAN	Q9Y6V0 homo sapien
9	651	8.7	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
10	629	8.4	6632	1 UNB9_GAEEL	Q01761 caenorhabdi
11	578.5	7.7	5120	1 PCLO_CHICK	Q9PU36 gallus gall
12	573	7.6	2812	1 ZAN_HUMAN	Q9Y933 homo sapien
13	559	7.4	875	1 FPI_MYTEO	Q25460 mytilus edu
14	551	7.3	620	1 EXTN_TOBAC	P13983 nicotiana t
15	551	7.3	1087	1 NPH_MOUSE	P19246 mus musculu
16	530.5	7.0	1162	1 TCNA_TRYCR	P23253 trypanosoma
17	530	7.0	865	1 CPN_DROME	Q02910 drosophila
18	518.5	6.9	872	1 FPI_MYTCO	Q25434 mytilus
19	517.5	6.9	831	1 NPI_RAT	Q16884 rattus norv
20	509.5	6.8	662	1 MUC1_XENLA	Q05049 xenopus lae
21	503.5	6.7	1970	1 RPB1_HUMAN	P24928 homo sapien
22	502	6.6	1020	1 NPH_HUMAN	P12365 homo sapien
23	498.5	6.6	1970	1 RPB1_MOUSE	P08775 mus musculu
24	493.5	6.6	467	1 RPB1_CRIGR	P11414 cricetus
25	490	6.5	2142	1 BAT2_HUMAN	Q48843 homo sapien
26	488.5	6.5	826	1 SSP2_PLAYO	P01443 plasmodium
27	478.5	6.4	857	1 APPI_SCHPO	Q9P768 schizosacch
28	475.5	6.3	267	1 EXTN_MAZE	P14918 zea mays (m
29	471.5	6.3	5376	1 ZAN_MOUSE	Q08799 mus musculu
30	469.5	6.2	1509	1 GSRI_HUMAN	Q9HMD4 homo sapien
31	468.5	6.2	634	1 HWP1_CANAL	P46593 candida alb
32	467	6.2	817	1 VRP1_YEAST	P37370 saccharomyc
33	454	6.0	797	1 VGLX_HSVB	P28968 equine herp

34	454	6.0	1161	1 DAN4_YEAST	P47179 saccharomyc
35	453.5	6.0	1794	1 YDC9_SCHPO	Q10172 schizosacch
36	450	6.0	3178	1 Y689_GAEEL	Q09696 caenorhabdi
37	448.5	6.0	670	1 VGS0_HSV1	Q00130 iccalturid h
38	446	5.9	1229	1 P121_HUMAN	Q09223 homo sapien
39	444.5	5.9	2774	1 MABA_RAT	P34926 rattus norv
40	442.5	5.9	751	1 PPI_MYGA	Q27409 mytilus gal
41	439.5	5.8	1083	1 T2D3_HUMAN	Q00268 homo sapien
42	436.5	5.8	1411	1 TCOF_HUMAN	Q13428 homo sapien
43	434.5	5.8	3256	1 K167_HUMAN	P46013 homo sapien
44	433.5	5.8	3164	1 TEGU_HSV1	P10220 herpes simp
45	432.5	5.7	439	1 XP2_XENLA	P17437 xenopus lae

ALIGNMENTS

RESULT 1
MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC 002817; Q14878; 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.B.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=9358717; PubMed=1885763;
RA Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC - FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCOS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC - SUBUNIT: MULTIMERIC.
CC - TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC - PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC - POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC - SIMILARITY: Contains 1 C-terminal cysteine knoc-1like (CTCK) domain.
CC - SIMILARITY: Contains 1 TIL (trypsin inhibitory-like) domain.
CC - SIMILARITY: Contains 2 VWFC domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; L1998; AAB55295.1; -				
DR	EMBL; M74027; AAB59875.1; -				
DR	EMBL; M94131; AAB59163.1; -				
DR	EMBL; M94132; AAB59164.1; -				
DR	PIR; A49963; A43932.				
DR	Genew; HGNC:7512; MUC2.				
DR	MIM; 158370; -				
DR	GO:0005803; C:secretory vesicle; TAS.				
DR	InterPro; IPR006208; Cys_knot.				
DR	InterPro; IPR006207; Cys_knot_C.				
DR	InterPro; IPR006209; EGF_like.				
DR	InterPro; IPR002919; TIL_Cyrtich.				
DR	InterPro; IPR001007; VWF_C.				
DR	InterPro; IPR001846; VWF_D.				
DR	Pfam; PF00007; Cys_knot; 1.				
DR	Pfam; PF01826; TIL; 1.				
DR	Pfam; PF00093; VWC; 1.				
DR	Pfam; PF00094; vwd; 4.				
DR	SMART; SM00214; VMC; 2.				
DR	SMART; SM00216; VMD; 4.				
DR	PROSITE; PS01185; CTCK_1; 1.				
DR	PROSITE; PS01225; CTCK_2; 1.				
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.				
DR	PROSITE; PS01208; VWF_C_1; 2.				
DR	PROSITE; PS50184; VWF_C_2; 2.				
KW	Glycoprotein; Repeat; Signal.				
FT	CHAIN	1	20	POTENTIAL.	
FT	FT	21	5179	MUCIN 2.	
FT	DOMAIN	1401	1747	APPROXIMATE REPEATS.	
FT	REPEAT	1401	1416	1.	
FT	REPEAT	1417	1432	2.	
FT	REPEAT	1433	1448	3.	
FT	REPEAT	1449	1464	4.	
FT	REPEAT	1465	1471	5.	
FT	REPEAT	1472	1478	6.	
FT	REPEAT	1479	1494	7A.	
FT	REPEAT	1495	1517	7B.	
FT	REPEAT	1518	1533	8A.	
FT	REPEAT	1534	1556	8B.	
FT	REPEAT	1557	1572	9A.	
FT	REPEAT	1573	1596	9B.	
FT	REPEAT	1597	1612	10A.	
FT	REPEAT	1613	1635	10B.	
FT	REPEAT	1636	1651	11A.	
FT	REPEAT	1652	1675	11B.	
FT	REPEAT	1676	1683	12.	
FT	REPEAT	1684	1699	13.	
FT	REPEAT	1700	1715	14.	
FT	REPEAT	1716	1731	15.	
FT	REPEAT	1732	1747	16.	
FT	DOMAIN	4815	4886	WFPC 1.	
FT	DOMAIN	4924	4991	WFPC 2.	
FT	DOMAIN	5075	5160	CTCK.	
FT	DISULFID	5075	5122	BY SIMILARITY.	
FT	DISULFID	5089	5136	BY SIMILARITY.	
FT	DISULFID	5098	5152	BY SIMILARITY.	
FT	DISULFID	5102	5154	BY SIMILARITY.	
FT	DISULFID	?	5159	BY SIMILARITY.	
FT	CARBOHYD	163	163	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	670	670	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	770	770	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	894	894	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1139	1139	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1154	1154	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1215	1215	N-LINKED (GLCNAC. . .)	(POTENTIAL).

FT	CARBOHYD	1230	1230	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	1246	1246	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	1787	1787	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	1820	1820	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4339	4339	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4351	4351	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4362	4362	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4373	4373	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4422	4422	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4438	4438	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4502	4502	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4616	4616	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4627	4627	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4752	4752	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4787	4787	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4881	4881	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4888	4888	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4955	4955	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	4970	4970	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	5019	5019	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	5038	5038	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CARBOHYD	5069	5069	N-LINKED	(GLCNAC .)	(POTENTIAL)
FT	CONFLICT	1351	1351	H -> L (IN REF. 3) .		
FT	CONFLICT	1412	1412	T -> S (IN REF. 3) .		
FT	CONFLICT	1449	1449	L -> P (IN REF. 3) .		
FT	CONFLICT	1504	1504	M -> T (IN REF. 3) .		
FT	CONFLICT	4192	4192	G -> S (IN REF. 2) .		
SQ	SEQUENCE	5179 AA;	540295 MW;	85CD7571FB9A5663 CRC64;		
Query Match		15.8%;	Score 1188;	DB 1;	Length 5179;	
Beech Local Similarity		30.9%;	Pred. No. 8.4e-38;			
Matches 405;		Conservative 87;	Mismatches 539;	Indels 279;	Gaps 54;	
Oy	4	KTLPIYLLLSVFIQOVSSODLSSC--AGRCG----	EGRSHDATCNCDCYNCOHMEC	56,		
Db	1169	KDRPIY-----EBDLKCKVTADKKCCGYEDNHPHGASVPPIETCKSCV-C	1213			
Oy	57	CPDFKRVCTAEILSCKGRCFESFERG-----RECDCAOQCKXDKCC-----PDYESFCAB	106			
Db	1214	TNSSQVCRRDE--EGKLNOTQGARCWEIICGNPTGEKHENICSIITRPSLTITFTT	1270			
Oy	107	VHNATSPS-SKAPRPSSGASQITKSTTKSRPKRPNKKTIKVIESELEIBEH--SVSEN	163			
Db	1271	ITLPPTTPISSTTTTTTTPTTSSIVLSTT--PK-----LCCLMSDWIMEDHSSSDSD	1320			
Oy	164	QESSSSSSSSSSSSTIKWIKSSKNUSAANRE-----LOKKLKVKONKNRTKKKPKPRPV	218			
Db	1321	GDRPPFGCGCAPEDI-EGRSYVDPHSLSEOHQAKQCVDSVGECIGNE-----	1368			
Oy	219	VDEAGSGLDN-GDFKYTTPDTSTTOHNKVSTSPK-IITAKPINRPSLRPNSDTSKET	275			
Db	1369	-DGNGNPFGICVDYKIRV-----NCCMPMDKITTPSPPTTSPPTTSPPTTTTLP	1419			
Oy	276	LTVNKEIVETKEITTTNNKOTSIDGKKTSAKETSISAKTOSIEKTSADLAPTSLAKLAPKA	335			
Db	1420	TTSPSPPT-----TTTPPPTTPSPPIITTTTTP-----LPTT-----TPSPPI	1458			
Oy	336	EITTKGPALTTPKEPT-----PTTPKEPASTTPKEPTTIKSAP-TTPKEP-----A	382			
Db	1459	STTTTPPTTPPTTSPPTTTPSPPTTTPPTTTPPTTTPPTTTPSPPTTPPTTPPASTTLP	1518			
Oy	383	PTTYSKAP-TTPKEPATTTKEBPATTPKEP-----APTTPKEPATTTKSAP-TTP	432			
Db	1519	PTTTPSPPTTTTTTPPTTTPSPPTTTPPIINRPTSTTLPPTTTPSPPTTTPPTTTPPTTTP	1578			
Oy	433	KEBPATTPKPAPITTPKEBPATTPKEPTPTTPKEBPATTPKEBPATTPKEP-----APT	485			
Db	1579	SPPTTTPSPPTITTTTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPT	1638			
Oy	486	APKBPATTPKEBPATTPKEBPATTPKEBPSTTPKEBPATTTKSAPITTPKEBPATTTKA	545			
Db	1639	TTSPSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSSPLTTPSPPTTMTT	1698			

```

Qy 546 P---TTPEKESPTTKEBAPPTTKEBAPPTPKKAPAPTPEKAPPTTKEBAPPTTKEBAP 602
Db 1699 PSEPTTSSSPI-TTTTSSSTTSPSPPTTMTTSPPTTSSPTTMTTLPPTTSSPLT 1757
Qy 603 TAPKEBAPPTP-----KETAPPTP----- 621
Db 1758 TTPPLPSIPTPTSPPTTPTTPTTVCPLCNMTGMLDSGKFNHFKPGDTELLIGVCGPGW 1817
Qy 622 -----KKLTP----- 626
Db 1818 ANNISGRATTPVPVPIGQLGQTVCDVSGVICKNEDQKRGVIRPAFCLINEINVOCEB 1877
Qy 627 --TTPKELAPTTPEKAPPTTPEELAPTTPEEPPTT--TPEBAPPTTAKAAPTPEKAPPT 683
Db 1878 CVTQPTTMTTTEENTPTTPTTITTTTTPPTTGTQPTTPTTITTTTTPPTTPTT 1937
Qy 684 TPEKP-----APTPEKAPPT--TPEKAPTTPKGTAPTTTKEBAPPTPKKAPKEAPPT 737
Db 1938 GTQPTTPTTPTTPTTPTTPTTGTQPTTPTTPTTPTTPTTPTTGTQPTTPTTPTTPTT 1997
Qy 738 TKEPTSTSDKAP--TTPKGTAPTTPEKAPPTTPEKAPPTTPEKAPPTTKEBAPPTPK 796
Db 1998 TPTPTPTGTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2056
Qy 797 PAKELAPTTTGTSTSDKAP--TTPKELAPTTPEKAPPTTPEKAPPTTPEKAPPTTPT 855
Db 2057 PTTTPTTPTTPTTPTTPTTGTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2115
Qy 856 EVSTPTTKEPTTHKSPESTPELSAETPKALENSPEK-----GVPT--TKTP 904
Db 2116 PPTPTGTQPTT---TPTTPTTPTTPTTPTTGTQPTTPTTPTTPTTPTTPTTPTTGT 2172
Qy 905 AATPEMTTAKOTTERDRL--TTP--ETTAAKMT---KETATTEKTESKIKTAP 956
Db 2173 TTTPTTPTTPTTPTTPTTGTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2232
Qy 957 TTVQVSTTQDPTTPTTKITLKTTLTTLAPKVTTL--KTIITTEIMNKEBETAKPDORATNSK 1015
Db 2233 PTP-PTGTPTTTP--ITT--TTTTPPTPTTGTQPTTPTTPTTPTTPTTPTTPTTGT 2287
Qy 1016 AATPKOKPTKAP--KPTSTKPKTTPRVAKKPTTPTPKMTSTMBELNP--TSHIABA 1071
Db 2288 TTTPTTPTTPTTPTTPTTGTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2347
Qy 1072 MGTTR--ENGPNKSLVENVKSEBAGABETPHMLARPHVMEBTP 1120
Db 2348 PTPGTGTPTTPTTPTTPTTPTTPTTGT--TQTP--TTTPTTPTTPTTPTT 2392

```

```

RESULT 2
SLP1_CLOTM STANDARD; PRT; 1664 AA.
AC 006852;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer
protein 1).
GN OLPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
NCBI_TaxID=1515;
RN 11)
RP SEQUENCE FROM N.A.
RX STRAIN=NCIB 10682;
RX MEDLINE=9320931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
"Organization of a Clostridium thermocellum gene cluster encoding the
cellulosomal scaffolding protein C1pA and a protein possibly involved
in attachment of the cellulosome to the cell surface.";
RT J. Bacteriol. 175:1891-1899(1993).
RC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.

```

```

CC -1- SUBCELLULAR LOCATION: Cell wall.
CC -1- SIMILARITY: Contains 4 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.1sb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; X67506; CAA47841.1; -.
DR PIR; T18262; T18262.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1664
FT DOMAIN 36 763
FT REPEAT 36 191
FT REPEAT 207 363
FT REPEAT 409 565
FT REPEAT 607 763
FT DOMAIN 771 1377
FT FT
FT DOMAIN 1378 1449
FT DOMAIN 1453 1494
FT DOMAIN 1495 1565
FT DOMAIN 1566 1625
FT DOMAIN 1626 1646
FT DOMAIN 1646 1664
SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA8FE74B CRC64;

Query Match 12.6%; Score 950; DB 1; Length 1664;
Best Local Similarity 31.0%; Pred. No. 2; 9e-29;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

Qy 327 VLAKPTP--KAETTTGPAALTTPKEPTTTPKESAPTTPEKAPPTTITKSAPTTKEBAPPT 385
Db 758 VVIQPAPIKASDEBIPDTSPDEPTPS-----DEPTPS--DEPTSPDEPTSD 804
Qy 386 TKSAPTTPEKAPPTTTPKEBAPTTTPKEA---PTTKEBAPTTTSAPTTPEKAPPTTPEK 441
Db 805 EPTPSBETPEBIPDTSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSBETPEBIPDTSP 864
Qy 442 KPAPTTPEKAPPTTKEPTTTPKESAPTTTPKEBAPPT--TPKEBAPPAKKAPDTTPEKAP 500
Db 865 SDEPTPSDEPTPS--DEPTPS--DEPTP--SDEPTSETEBEBIPDTSPDEPTPSDEPTP 919
Qy 501 TTPKEBAP--PTTKEBSPPT--TPKEBAPTTTTS--APTTPKEBAPPT---TTKSAPTTPEK 553
Db 920 SDEPTPSDEPTPSDEPTSETEBEBIPDTSPDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
Qy 554 ----PTTKEBAPPT--TPKEBAPTTTPKAPPTTTPKEBAPTTTPKEBAPTTTTPKAPPAK 608
Db 980 PSDEPTPSBETPEBIPDTSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1039
Qy 609 APT---TPKETAPTTPKLTPTTPEKLAPTTPEKAPPT---TPBELAPTTPEBPTT--TP 661
Db 1040 TSDSPETSEB-----TPBEBIPDTSPDEPTPSDEPTPSDEPTPSD--EPTPSDEPTSE 1094
Qy 662 EEPATTTPKAAPTTPKEBAPTTTPKEA--PTTPEBAPPT--TPKETAPTTTPGTAPTTT 718
Db 1095 EEPITPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1154
Qy 719 EAPTTTPKAPKELAPTTTKEPT--STSDKAPPTTTPGTAPTTTTPKEBAPTTTPKEBAPPT 777
Db 1155 EPTPS--DEPTPSD--EPTPSDEPTSETEBEBIPDTSPDEPTPSDEPTPS--DEPTPS- 1208
Qy 778 PKGTAPTTTKEBAPPT--TPKAPKELAPTTTGTSTSDKAPPT---TPKETAPTTPEK 833
Db 1209 ---DEPTPSDEPTSETEBEBIP-----PDTSPDEPTPSDEPTPSD--EPTPSDE 1253

```


DB 451 GSVANVTIRVAFATEK--PALIYSIELVYNTG-----ATLRVPIANVTRSGIR 500
 QY 990 ----KTTTTEIMKPEETAKKORATSKATP--PKPKPTKAPKPKSTTKPT 1039
 DB 501 CPGLTYGTPTDIAGP-----TQIDATYTMNKIAGVRLMGAGNKKPKT 545

RESULT 5
 PCLO RAT STANDARD; PRT; 5085 AA.
 ID PCLO RAT
 AC Q9JUS6; Q9JUL1; PRT; 5085 AA.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
 GN PCLO.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
 RX MEDLINE=20170257; PubMed=10707984;
 RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
 RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
 RT "Piccolo, a presynaptic zinc finger protein structurally related to
 RT bassoon.",
 RL Neuron 25:203-214(2000).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RN CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668, ASP-4674;
 RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
 RP ALA-4694.
 RX MEDLINE=21181819; PubMed=11285225;
 RA Geber S.H., Garcia J., Rizo J., Suedhof T.C.;
 RT "An unusual C(2)-domain in the active-zone protein piccolo:
 RT implications for Ca(2+) regulation of neurotransmitter release.",
 RL EMBO J. 20:1605-1619(2001).
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (by synaptic).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin.
 CC -1- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
 CC junctions.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9JUS6-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9JUS6-2; Sequence=VSP_003930, VSP_003931;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DRH domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF138789; AAF07822.2; -;
 DR EMBL; AF275534; AAF63196.1; -;
 DR HSBP; P04410; I425.
 DR GO; GO:0045202; C:synaptic junction; IDA.
 DR GO; GO:0005509; F:calcium ion binding activity; IDA.
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; IDA.

DR GO; GO:0005522; F:profilin binding activity; ISS.
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 1.
 DR PROSITE; PS50004; C2_DOMAIN_2; 2.
 DR PROSITE; PS50106; PDZ; 1.
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
 KW Repeat; Alternative splicing.
 FT DOMAIN 372 491
 FT 12 X 10 AA TANDEM APPROXIMATE REPEATS OF
 FT P-A-K-P-Q-P-Q-P-X.
 FT C4-TYPE (POTENTIAL).
 FT C4-TYPE (POTENTIAL).
 FT POLY-PRO.
 FT PDZ.
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT TKPTN -> SKRRK (in isoform 2).
 FT /Frid-VSP_003930.
 FT Missing (in isoform 2).
 FT /Frid-VSP_003931.
 FT D->A: COMPLETE LOSS OF CALCIUM-BINDING
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID
 FT BINDING ACTIVITY.
 FT V->S: SMALL INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT M->S: INCREASED AFFINITY FOR CALCIUM.
 FT VV->SS: 10-FOLD INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT ON->AA: MODERATE INCREASE IN AFFINITY FOR
 FT CALCIUM.
 FT A->S: NO EFFECT ON CALCIUM-BINDING
 FT ACTIVITY.
 SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
 Query Match 9.3%; Score 703.5; DB 1; Length 5085;
 Best Local Similarity 24.1%; Pred. No. 1.3e-19;
 Matches 313; Conservative 118; Mismatches 437; Indels 431; Gaps 60;
 QY 113 PPSKKAPPEGASOTIKSTTKRSP-KPP-----NKKTKKYIASEE----- 153
 DB 68 PPAABSPSMNRKKELDSSQAPQPGKPPDGRPTGSLSKRTTDFRSQKIPGSRPS 127
 QY 154 ---ITTEHSEVSENGESSSSSS-----SSSSSTTWIKS-----SKNSANRELQK 196
 DB 128 TISLSEKSRDFFBEYKSSWMPGFDVNPVLSAVSVVKNFNPFDLISDEAGQEEETK 187
 QY 197 KLVKDKNKKRNTKKKPPKPPVVDKAGSLDNGPKVTPPTSTTQHNKSVTSKITTK 256
 DB 188 KQKV- -QKEGKSEGNKAPPL-----Q 208
 QY 257 PINRPSLPNDSKETSILVNKETTVERKETTNNKQSTDSGEKTSKAKETQSIETK 316
 DB 209 QPSKP-IPKQGGVKEVIQDSSPKVSSQAKVKPQABGTGK-----PSQSPAQOT 261
 QY 317 SAKDLAPTSKVLAP-TPKATTTKGPALTPPKPEPTTPPKPASTTPKEPTPTTIKSAP 375
 DB 262 PAQASPGKPAVQPGSAKATVQGPDA-KSPA-PA-GTGKSPAQPAKTRGQQAAGLEKT 319
 QY 376 TTPKEPAPTTTKKAPTP-----KEPA-PTTKEPAPTPPKAPAP-TTTKEPAP 422
 DB 320 SSSQPGP-----KSLAQPTGKHKFPLGPKVKSAPQAPGAKHPA-----QDPGPTAKVGP 372

Query Match	8.8%;	Score 665;	DB 1;	Length 5038;
Best Local Similarity	24.5%;	Pred. No. 3.7e-18;		
Matches 300;	Conservative 129;	Mismatches 449;	Indels 348;	Gaps 60;

Qy	110	TSPPSSKAPPPSAOSOTIKSTRSPK--PBNKKTKKVISEIETHEHSVENOS-	166
Db	93	EKRPDPBR-PPOHGLSKSRITDTIRSEOKLPGRBPSTILSKESSRD--FKXEYSS	148
Qy	167	-----SSSSSSSSSTTWIKS---SKUSANRELQKLV--KONKNRTKKKPTP	214
Db	149	MMGPFSSVNPPLSAVSSVYNKNPNPDLLISDSEAVQEEFTTKQVAKQOGKSEGITKPSL	208
Qy	215	KPVPVDEAGSGLDNDPFVTTPTDSTTQHNKVS PKITLAKIPNRPBSPENDSTSEKT	274
Db	209	QOP-----SPKL-----IPKQOGRKEV	226
Qy	275	SLTVNKEITVEKETTTNNKOTSDGEKETSASKEOTSIEKTSADLAPTSKVLAFTPK	334
Db	227	I-----PDI PSKVS SGOAKET--KQAPGPAKRSQGSQAPQTPAQOAKPVAQ	272
Qy	335	AETTTGPA LTTKEPTPT--TPKEPAST--TPKEPTPTTIS-----APTTPKEPA PT	384
Db	273	----OPGPAKTVQOPGPAKSPAQAGGKSPACQ--PVTAKPRAOAGLEKTSLOOPGPK	327
Qy	385	TTKSAPTTPKEBAPTTTKEBA--PTTPKEBAPTTTKEBAPTTTKEBAPTTTKEBAP	442
Db	328	SLAQTPGGGKVP--PQPAKSPAQOFGTALPRA----QOQEOYASKVPEFTKPAQLS--G	380
Qy	443	PAPTTTKEBAPT--TPKEPTPTTPKEBAPTTTKEBAPTTTKEBAPPAKKBAPTTTKEBAP	500
Db	381	PGKTPAQOGPFTKPSPOQPIPAKPOQPVATPK--QOQCAPAKPFQ-----POHPTP	431
Qy	501	TTTPKEBAPTTTK--BSPTTPKEBAPTTTKSATTTTKEBAPTTTKEBAPTTTKEBSPTTTKE	559
Db	432	AKPOQOQTPAKPOQOFTPAKPOQHQFGLKPSAQOBSKSSISQTVTGAPLOAPPTSAQO	491
Qy	560	-PAPTTTKEBAPT-----TPKK-----PAPTTTKE---	583
Db	492	APAGLSTICPLCMTTELLHTPEKAPANFTCEGOSTVCSLGNPNPHLLEIKEWLCL	551
Qy	584	-----PAPTTTKEBAPTTTKKPAPTAPEBAPT--TPKETAPTTPKKLPT	627
Db	552	NCOMORALGELAIAPSSP-QPTPAASVQOPATASKSPVPSQOASPKELPS--KQDSBK	608
Qy	628	TPKCLAPTTPEKBAPTTPEBELAPTTPEBPTPTTPEP--APTTPKAANPNTPKEBAPPTP	685
Db	609	APES-----KKPBPVAKQ--PLHGGTPTAPQOPVAEALPBPAPPKKRSAALEPOA	658
Qy	686	KEP-APTTPKEBAPPTTPEKETAPTTKGTAPTTKEBAPPTTKEBAPKAPKELAPTTTKEPTST	744
Db	659	KAPVADVPEPKQPKT---ETLDSBSSAAIS--KPA LLSQVQAO--AQVTTAPPLKT	710
Qy	745	TSDXPAPTTPKGTAPTP-----KEBAPT-----	768

Dd		711	DSATSGSFPPTGGDTLTPDLSKAMPBPASDSKIIVSHGPFSSSKDPVQKEEPKAQTkv	770
Oy		769	TPK---BPAP---TTPKGIAPT--LKRPAPTPPKPAKPELAFTTGTSPSTSDKRAPt	821
Dd		771	TPKEDTRFVPKGSPPTPGIRPTTGQAATPQSQQPPKPEQGRRRFLNLGGIA-DAPKSQPt	829
Oy		822	TPKETAP-----TTPKBPAPTPPKKAPAPTPTEPTEPPTSEVSTPTTt	863
Dd		830	TPQETLVGKLFGFCASLIFSQASNLISITAGQAAPHPIQGPAAPSQAPFPSoT-----	881
Oy		864	KEPTTIHKSPEDESITBELSAEPTPALNSPKPEGVPT-----TKTPAA-----TKPEMTTt	914
Dd		882	-----LAAQPSPKSTQGHSPASAPAKTTAVKKETKGAENLEAKPAQApt	925
Oy		915	AKDKTERDLATTYETTTLAAPKMKTELATTT---EKTTESK----IATTTQVYSTTDQ	967
Dd		926	V--KKAEKDCKHPGKGSKPPEPEPEKAVLAQKDKTTTKRKPCAPCLRTIELANGS----	980
Oy		968	TTPr-----KITLLKTTLTAAPKVt-----TYKKTIP--TTEIMNKPEETAKP	1007
Dd		981	PNNNTCTECKNQCNLCGNPNPLHLTEIQEWLCLNCTORALISGQLGDWDKMPASSgP	1040
Oy		1008	KDRATNSKATTPKQKQPT---KAPKKPTSTK--KPTMPPVRVKPTTPPr-----	1052
Dd		1041	KASVPAPEBPPQKPTTLAAHGKKGKETEYKATEKQIQIEKENPTSIEKTPPAVAIDQkL	1100
Oy		1053	-----RKMTSTMELNPFSRIAEMLIQTTTRNQTNsk-----LVEVNPKSED-----	1096
Dd		1101	ESEVTSLIVSLVEPKKPSER--EKALPADKKEKKPPAALAPPLEKKPRIPDDQKLPPDA	1158
Oy		1097	-AGABGETPHMLLRPHVMPEVTPt	1120
Dd		1159	KPASASEGEKRDLKAHVQIPEEGP	1183

RESULT 7

MUSB_HUMAN		STANDARD;	PRT; 5703 AA.
ID	"MUSB_HUMAN		
AC	Q95HC8; Q00447; Q00573; Q14985; Q15494; Q95291; Q95451; Q14881;		
AC	Q95552; Q9UE28;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High		
DE	molecular weight salivary mucin MGI) (Sublingual gland mucin).		
DN	MUC5B OR MUC5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RN	SEQUENCE OF 1-1594 FROM N.A.		
RA	Chen Y., Di Y.P., Wu R.;		
RL	"Molecular cloning of the amino-terminal and 5'-flanking region of the		
RL	human MUC5B mucin gene.";		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RN	SEQUENCE OF 1-1325 FROM N.A.		
RX	MEDLINE=99009274; PubMed=9790959;		
RA	Offner G.D., Nunes D.P., Keates A.C., Afhdal N.H., Troxler R.F.;		
RT	"The amino-terminal sequence of MUC5B contains conserved		
RT	multifunctional D domains: implications for tissue-specific mucin		
RT	functions.";		
RL	Biochem. Biophys. Res. Commun. 251:350-355(1998).		
RN	[3]		
RE	SEQUENCE OF 40-1324 FROM N.A.		
RX	MEDLINE=99023933; PubMed=9804771;		
RA	Deesey J.-L., Bulsine M.P., Porchet N., Aubert J.-P., Laine A.;		
RT	"Genomic organization of the human mucin gene MUC5B: cDNA and genomic		
RT	sequences upstream of the large central exon.";		
RL	J. Biol. Chem. 273:30157-30164(1998).		
RN	[4]		
RE	SEQUENCE OF 1326-4895 FROM N.A.		

FT CONFLICT 104 104 S -> C (IN REF. 2).
 FT CONFLICT 142 142 E -> K (IN REF. 1).
 FT CONFLICT 225 225 R -> S (IN REF. 2).
 FT CONFLICT 330 331 PL -> T (IN REF. 2).
 FT CONFLICT 337 337 E -> N (IN REF. 2).
 FT CONFLICT 356 356 E -> K (IN REF. 2).
 FT CONFLICT 362 362 G -> R (IN REF. 2).
 FT CONFLICT 369 369 MISSING (IN REF. 2 AND 3).
 FT CONFLICT 374 374 D -> N (IN REF. 2).
 FT CONFLICT 393 394 RT -> TR (IN REF. 2).
 FT CONFLICT 468 468 RK -> GR (IN REF. 2).
 FT CONFLICT 512 512 L -> P (IN REF. 2).
 FT CONFLICT 585 587 GAA -> AH (IN REF. 3).
 FT CONFLICT 601 601 A -> S (IN REF. 3).
 FT CONFLICT 628 629 DP -> RS (IN REF. 2).
 FT CONFLICT 633 629 F -> L (IN REF. 2).
 FT CONFLICT 676 676 A -> P (IN REF. 3).
 FT CONFLICT 701 701 R -> P (IN REF. 3).

Query Match 8.8%; Score 660; DB 1; Length 5703;

Best Local Similarity 22.0%; Pred. No. 6.3e-18; Matches 394; Conservative 120; Mismatches 579; Indels 698; Gaps 71;

Qy 54 MECCDFRVCYTAELSCRGCFSEFREGCCDCAQCKKDYKCCPDYE--SFCALV--NRP 110
 Db 2906 VECSLDFGLVC-----RMR-----QVGRF--KMCFNVEIRPFCNHYGHC 2944
 Qy 111 TSPSSKAPPPS-----GASQITKSTKSPKPKKKTKVIE 150
 Db 2945 STPATSSATPSSSTPGTWTILTEQTTAATTATTTATSTAISSSTPGTADPP-----KYL 2998
 Qy 151 SEETHEHVSSENGESS-----SSSSSSSSSTIKKISS-----KNSANR--- 192
 Db 2999 SQATTPATSSKAISSSSPRTATTLPVLTSTATSTATSPPIPSSTLGTGTSQNRPH 3058
 Qy 193 -----ELQKKLVKDNKKRKKKPPKPP-----VVDE-----AGS 224
 Db 3059 PMATMSTHPSSTPETHSTVLTATTKATTRATSMSTPSSSTPGTWTILTELTAAATTA 3118
 Qy 225 GLDNG-----DFKVTTPDSTQHN-----KVSSTPKITTAKEPI 258
 Db 3119 ALPHGTSSSTPGTWTILTEPSTTAIVPTGTSSTAASTRAATAGTLKVLTS--TATTPVI 3177
 Qy 259 NPR-----PSLPNSDTSKETSILVNKETTVEF-----KETTTNKOQSDG 300
 Db 3178 SSRATPSSSPGTATAPALRSTATTPTATSVTAIPSSSLGTAMWRLSQTTPTATMSTAT 3237
 Qy 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKETPKAETTK-----GPA 343
 Db 3238 PSSSTPEVHTSTVLTATTTATTRIGSVAPSSSTPGTAHTTKVPTTTTGTATPSSSPGTA 3297
 Qy 344 LTPP--KEPTPT--TPKE-----PASTPKEPTP 368
 Db 3298 LTPPWTISTTTPTTRGSTVPSSIPGTHATATVLTATTTTAVANGSMATPSSSTQISGTP 3357
 Qy 369 TTISAPTTKEPAPTTPKAP--TPKEPAPTT--KEPAP--TPKEPAPTTPKEPA 421
 Db 3358 PSLTTATTTATATGSGTNPSSSTPGTTPIPVLTATTTATTPAATSTVPSSALGTHHPV 3417
 Qy 422 PTTTKSA-----PTTP-----KEPAPTTPKAPPT--PKEP 451
 Db 3418 PNTIATHTGRSLPSSPHITVPTAMTSATSGILGTHIETPSTGSHTPAATGTTGTQSTP 3477
 Qy 452 APPTPKAPTTPKEPAPTTPKEPAPT--TPKEPAPTAPKAPAPTTPKEPAPT--PKEPAP 508
 Db 3478 ALSSPHSSRTTSSPSPGTTTTPGHTRGTSRTTATATPSKTRSTLLPSSSPTSAPIITVV 3537
 Qy 509 TTTKEPS----- 515
 Db 3538 TTGCEPQCAMSEMLDYSYMPGPGSGDFDTYSNIRAGAVCEOPLEGECRAQAOPGVPL 3597
 Qy 516 -----PTTPKEPAPTTPKAP 531

Db 3598 RELGQVVECSLDPGVLCNRRBQVKKFKMCFNYEIRFCCNHYGHCSTPATSSSTP 3657
 Qy 532 TTT--KEPAPTTPSAPTTPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 589
 Db 3658 GTTILATLTATTTATTTESGTATPSSSTQGPAPTPH--VSTTATTPVITSSK--ATP 3711
 Qy 590 KEPAPTTPKAP-----TPAPKEPA-----PTTPKEPAPTTPKAP 626
 Db 3712 FSSPGTATAPALMSTATTPTATSTFATPSSSLGTTWRLSQTTTPMATMTATPSSSPE 3771
 Qy 627 -----TPPEKAPPTTPKEPAPT--PEELAPTTPEED-- 656
 Db 3772 TVHTSTVLTATTTATTTAGATSVATPSSSTPGTAHTTKVPTTTTGTVPSSPGTARPPV 3831
 Qy 657 ---TPTPPEAPPTTPKAPAAANTPEKAP--TPKEPAPT--TPKEAPPT--TPKE--T 704
 Db 3832 WISTTTPPTSSGTVTTPSSIPGTHPTVLTTPQPAVATGSMATPSSSTQSGTPPSLIT 3891
 Qy 705 APPTPKAPTTPKAPAPTTPKAPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 753
 Db 3892 TATTTTATGTT--NPSSSTPGTTPIPBELTTATTPAATSSSTVPSSALGTHHPVNT 3949
 Qy 754 PKGT-----APTP-----KEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTP 787
 Db 3950 TATTHGRSLSPSSPHITVPTAMTSATSGILGTHIETPSTGSHTPAATGTTTSTPALS 4009
 Qy 788 EPAP--TPPKAPKAPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 834
 Db 4010 SPHSSRTTSSPSP--GTTTPGHTTATSTRTTATATPSKTRSTLLPSPQTSAPITVV 4065
 Qy 835 ----- 834
 Db 4066 VTGCEPQCAMSEMLDYSYMPGPGSGDFDTYSNIRAGAVCEOPLEGECRAQAOPGV 4125
 Qy 835 -----APTPKAPAPTTPETP 850
 Db 4126 LGEIGQVVECSLDPGVLCNRRBQVKKFKMCFNYEIRFCCNHYGHCSTPATSSSTP 4185
 Qy 851 PPTT--SEVSTPTTPKEPTTHKSPPDESTPELSAETPKALENSPKPEPGVPTTK-- 902
 Db 4186 PGTWTILTELTATTTATTTASGSTATP--SSTP--GTAPPKULTSPATTPATSSKATSSS 4242
 Qy 903 -----TPAATK-----PEMTTAKDKTTERDLRTTPEPT 931
 Db 4243 SPRTATTLPVLTSTATSTATSVTPIPSSSLGTTGTLPEQTTTVVATMSTHPSSTPETH 4302
 Qy 932 -TAAPKMTKETATTEKT-----TESKITATTTQVTS--TTQDTP-- 970
 Db 4303 HTSTVLTATTKATTRATSSSTPSSSTPGTWTILTELTATTTATTTAGTATPSSSTPGTWTIL 4362
 Qy 971 FKITTLKTTTLAPKVTTT--KTTITTEIANKPEETAK-----PKDATNSKATTPKAPQ- 1022
 Db 4363 TELTTTATTTASSTGSLTSLSTPGTWTILTEPSTTAIVPTGTSSTAASTAAATGAPHV 4422
 Qy 1023 -----KPTKAPKKTSTKPKKT--MPRVKPKTTPPRKMT----- 1056
 Db 4423 STTATTPVSSSKATTPSSSPGTATAPALMSTATTPTATSTFATPSSSLGTTWRLSQTT 4482
 Qy 1057 -----STWPELNPTRIAEAMLQTT-----TRNQPNSKLVEVNPKESEDAG- 1098
 Db 4483 TPTATMTATPSSSPETVHTSTVLTATTTATTTAGATSVATPSSSPGTATHTTKVPPTTTTGF 4542
 Qy 1099 -GAGETPHMLLRHVMEVTPMDVLPRVNOGIIINPMLSEBETNICNKPVDGLTTL 1157
 Db 4543 TAPSSSPGATLPPVWISTTTPPTTTP--TSSGTVTPSSSLPGTHHTARVLTITTTTV 4600
 Qy 1158 RNLGLVAFRGHYFWMLSPPSPSPARITEVWIGPSIDIVTFRCNCEGKT 1208
 Db 4601 ATGSM-----ATPSSS--TQTSIPSLTITATTTATGST 4634

RESULT 8
 PCLO_HUMAN

ID PCLO_HUMAN STANDARD; PRT; 5147 AA.
 AC Q9Y6V0; Q43373; Q60305; Q9BVC8; Q9UIV2; Q9Y6U9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Acronin) (fragments).
 OS PCLO OR AC2 OR KIAA0559.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE OF 1-759 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99439764; PubMed=10508862;
 RA Wang X., Kibbechull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
 Kitzmann M.W.;
 RT "Acronin, a 550-kD putative scaffolding protein of presynaptic active
 zones, shares homology regions with rim and bassoon and binds
 profilin.";
 RT J. Cell Biol. 147:151-162(1999).
 RN (2)
 RP SEQUENCE OF 552-4404 FROM N.A.
 RC Kraemer J., Woliam C., Wohldmann P., McGrane B.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 3619-5147 FROM N.A. (ISOPFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628591;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 RN (4)
 RP SEQUENCE OF 4405-4439 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Struhsberg R.L., Felingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Schenken C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Burow K.H., Scheffer C.F., Bhat N.K.,
 Hopline R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Frange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
 Bosak S.A., McEwen P.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Gilwood J., Schmutz J., Myers R.M.,
 Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scheraga A., Schein J.B., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RL (5)
 RP SEQUENCE OF 4405-5147 FROM N.A.
 RA Kalicki J., Billicot G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
 CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;

CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipide. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: Y19188; CAB60727.1; -;
 CC EMBL: AC004886; AAD20936.1; -;
 CC EMBL: AC004886; AAD21789.1; -;
 CC EMBL: AB011131; BAA25485.1; -;
 CC EMBL: BC001304; AAH01304.1; -;
 CC EMBL: AC004082; AAB97937.1; -;
 CC PIR: T00634; T00634.
 CC HSSP: P04410; 1A25.
 CC GeneW: HGNC:13406; PCLO.
 CC MIM: 604918; -;
 CC GO: GO:0005856; C:cytoskeleton; NAS.
 CC GO: GO:0045202; C:synaptic junction; ISS.
 CC GO: GO:0005509; F:calcium ion binding activity; ISS.
 CC GO: GO:0005544; F:calcium-dependent phospholipid binding; ISS.
 CC GO: GO:0005522; F:profilin binding activity; ISS.
 CC GO: GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 CC GO: GO:0016080; P:synaptic vesicle targeting; ISS.
 CC InterPro: IPR000008; C2.
 CC InterPro: IPR015655; Synaptotagmin.
 CC PRINTS: PR00360; C2DOMAIN.
 CC PRINTS: PR00399; SYNAPTOTAGMIN.
 CC SMART: SM00239; C2; 2.
 CC PROSITE: PS00499; C2 DOMAIN 1; 1.
 CC PROSITE: PS00004; C2 DOMAIN 2; 2.
 CC KX Calcium/phospholipid-binding; zinc; Metal-binding; zinc-finger;
 CC Repeat; Alternative splicing.
 CC FT NON_TER 1
 CC FT DOMAIN 1
 CC FT 400 465
 CC FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
 CC P-A-K-P-Q-P-Q-P-X.
 CC FT ZN_FING 499 523
 CC FT C4-TYPE (POTENTIAL).
 CC FT ZN_FING 969 992
 CC FT C4-TYPE (POTENTIAL).
 CC FT NON_CONS 1010 1011
 CC FT POLY-PRO.
 CC FT DOMAIN 2300 2325
 CC FT PDZ.
 CC FT DOMAIN 4391 4432
 CC FT C2 DOMAIN 1.
 CC FT DOMAIN 4544 4633
 CC FT C2 DOMAIN 2.
 CC FT DOMAIN 5031 5121
 CC FT S -> SCNGICLRVGGKEIRGSGEIGAYIAKLPGSAA
 CC VARSPLIC 4404 4404
 CC QTKLMES (in isoform 2).
 CC FT FTId=VSP_003923.
 CC VARSPLIC 4534 4534
 CC K -> KPTDGRKVSHPRTGEIQ (in isoform 2).
 CC FT FTId=VSP_003924.
 CC VARSPLIC 4576 4576
 CC G -> GQWVIVONAS (in isoform 2).
 CC FT FTId=VSP_003925.
 CC VARSPLIC 4757 4761
 CC TAHS -> EGRK (in isoform 2).
 CC FT FTId=VSP_003926.
 CC VARSPLIC 4762 5147
 CC Missing (in isoform 2).
 CC FT FTId=VSP_003927.
 CC SQ SEQUENCE 5147 AA; 563537 MW; CDS08490498CD3C CRC64;
 CC Query Match 8.8%; Score 659.5; DB 1; Length 5147;
 CC Best Local Similarity 23.7%; Pred. No. 6e-18;
 CC Matches 302; Conservative 143; Mismatches 494; Indels 337; Gaps 60;
 CC Oy 113 PPSKKAAPPSSG-----ASQTIKSTTKSPKPPNKKTKVIESEITEHSVS 161

```

Db      56 PPKSGRRPDDGRAPQBLGSKSRITDTFRSROKLPGRSPSTILSKESKSRDLDKEHSS 115
Qy      162 ENOE-SSSSSSSSSSSTIMTKS---SKNSANRELOKLVKDNKKNRTK---KXP-212
Db      116 MPPGLSEVNLAISAVSSVANKFNFPDLISDEASQSEFTTKQKVQKQKQEGEIIKPL 175
Qy      213 --TKRPVVDAGSLDNGDFKVTTPDSTTQHNVSTSPKITTAKPINRPSLBPBNDT 270
Db      176 QQQPFPKIPKQGGPERDPLQODGTPKSISQOPKIKSQPP-GTKPIQ-GTGTPTQTDH 233
Qy      271 SKETSLTVNKEETVETKETTTTNNKQTSDEKKTISAKETOSIEKTSADLAPTSKVLAK 330
Db      234 AK---LPLQRD-----ASRPQTQADIVRESVKSPLSPSPSK 267
Qy      331 PTPKAETTKGPAITTPKEPTTTPKEBAPTTPKEPTT---ISAPITPKBPA---PTT 386
Db      268 PPIQOPTGKPPAQQPGHEKSGQPGPAKP---PAQPSGLTKPPLAQOQGTIVRPPVOPPTT 323
Qy      387 KSAPITPKBPAITTPKEBAPTTPKEBAPTTPKSAPTTPKKEBAPTTPKKEBAPT 446
Db      324 K-PPAQPAGPA---KPPAQOTGSE--KPSBQEPFKLADPPGVGKTPA---QQPQPA 372
Qy      447 TPKEBAPTTPK---EPPTTP--KEBAPTTPKEBAPTTPKEBAPTTPKKEBAPT 501
Db      373 KPPQOVGTGPKPLAQOQPLQSPAKAPG-TKTPAQTKPSPQOPSTKTP---PQQGPBA 427
Qy      502 TPKEBAPTTPKPS---PTTPKEBAPTTPKSAPTTPKEBAPTTPKKEBAPT 557
Db      428 KPSPOQPGSTKPPSQPGSAKPSAQOQTPKPVSGTGFGRPLPPTVSPSAK 487
Qy      558 KEBAPTTPKEBAP---TT-----PKK-----PAPTTPKE---583
Db      488 QPSQGLBKITICPLNTTELLHVEKANTCTECQTVCSLGFENPHLTKAKELC 547
Qy      584 -----PAPTTPK---EPAPTTPKAPAPTTPKKEBAPTTPKKEBAPTTPK 626
Db      548 LNCQMKRALGGDLAVPSSPQPKLTAIVTTTSAVSKSPQOQSPKKA---APKQDLS 605
Qy      627 TTPKEBAPTTPKEBAPT-----TPEELAPTTPKEBAPT-----PTTEBAPTTPKKAAP- 674
Db      606 KAPB-----PKKEPPLVKQPTLNGSPSAKAKQPEADSLSKPAPKPSVSEDDKAPV 659
Qy      675 --NTPKEB---APTTP---PKBAPTTPKEBAP-----PTTPKEBAPTTPKSTATT 716
Db      660 ADDPKQPKMVKPITDLVSSSSATTKPDIPSSKVSQABEKTTPKLDKSAKPSQSPPT 719
Qy      717 -----LKEBAPTTPKPAKELAPTTPKEPTSTSDKPAAP-----751
Db      720 GEKTPPDSKAIIPRASDSKIIISHGPSSESK-GQKQVDPVQKKEPKAQTKMSPKDA 778
Qy      752 -TTPGGLAPTTPKEBAPTTPKEBAPTTPKGAAP-----TTLKEBAPTTPK 796
Db      779 KPMKGS---PTPPGPRPTA-GQTVPTPOQSPKQOSRRFSINLGSIDAVKSQPTTPOE 835
Qy      797 PAKKEL-----APTTPKGSTSTSDKPAAPTTPKEBAPTTPKPAAP- 834
Db      836 TVTGKLFQFGASISQASNLISTAGQPGPHSOSGGAPMKAQAPASQ---PTSQGPBK 891
Qy      835 -----APTTPKPAAPTTPKEPTTSEV-----STPTTPKEPTTIHKSDESTBELSAP- 884
Db      892 STGQAPAPAKSIPPKETKAPAAKLEPKAQAQAPTVKRTETKEKPPRIKOSKSLTAPQ 951
Qy      885 ---TPKALENSPK-EPGVPTTKP---AATPEKTTAKDKT-----919
Db      952 KAVLPKLEKSPKESTCPLCTEINISKDPENNTCTEKNQVCNLGFPPTPLETEN 1011
Qy      920 ---TER-----DRTPEPTTA---APMKTETATTEKTKT---ESKITATTTQVTSSTT 965
Db      1012 CQTAQASIGQGLDKMPPADSGPAPASPMVPTSSSSQKTAVPPQVKLVKQEQGVKTEA 1071
Qy      966 QDTTPFKI-TTLKTTTLAPKVTTTKTITTEINMKPEETAKPKORATNSKATTPKPKP 1024

```

```

Db      1072 EKVLIEKVEKTELMEKIPMWTTDQ-----KQESKLEKDKASALOQEKPLPEEK 1121
Qy      1025 TKAPKKTSTSTKPKETMRVAKPKTTPPRK-----TSTMP-----LNPFSRIAEAMQ 1074
Db      1122 KLIPBEERKIREEK-KPLBEKPKTPEDKLLPBAKTSAPBEOGHDLKQVQIAEKELE 1180
Qy      1075 TTPRNOGTPNSKIVENVKPSDAGAGETPFHMLLRPHVFMPEVTMPMDYLPRVNOGII 1134
Db      1181 GRVAPKTVQBEK-----QPTMEGLPSG-TPOSLEPKD---DKTKTIKQPPQPP-----1227
Qy      1135 INPMLSDETINCKGP 1150
Db      1228 -----CTAKP 1232

RESULT 9
MUC1 HUMAN
ID MUC1 HUMAN STANDARD; PRT; 1255 AA.
AC P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BX44; Q9UE75; Q9UE76; Q9UOL1; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H2JAG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Bacteria S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=90202794; PubMed=2318825;
RA Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wareschney D.H., Harevuani M., Tsarfacy I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kocbek P., Weiss I., Lache R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";

```


RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RN SEQUENCE FROM N.A.
RP TISSUE-Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Harevent M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RT Zrlan S., Weiss M., Green S., Lathé R., Keydar I., Wreschner D.H.;
RA "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue.";
RL Biochem. 189:475-486(1990).
RN [7]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP MEDLINE=9103045; PubMed=1688329;
RX Tsarfaty I., Harevent M., Horev J., Zaretsky J., Weiss M.,
RA Tsarfaty J.M., Garnier J.M., Lathé R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RN SEQUENCE FROM N.A. (ISOFORM 7).
RP MEDLINE=95010060; PubMed=7925397;
RX Zrlan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]
RN SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
RP MEDLINE=97355747; PubMed=9212228;
RX Coesterkamp H.M., Scheinert L., Stefanova M.C., Lloyd K.O.,
RA Pineda C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/2).";
RL Int. J. Cancer 72:87-94(1997).
RN [10]
RN SEQUENCE FROM N.A. (ISOFORM 7).
RP Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/X.";
RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RN SEQUENCE FROM N.A. (ISOFORM 9).
RP TISSUE-Epithelial cancer;
RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line.";
RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RN PARTIAL SEQUENCE FROM N.A.
RP MEDLINE=88330762; PubMed=347635;
RX Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rochard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [13]
RN SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).
RP MEDLINE=9008473; PubMed=2597151;
RX Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DP3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [14]
RN SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
RP TISSUE-Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Properative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [15]

RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE-Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RT Lee L.N., Luh K.T., Wu C.W.;
RA "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [16]
RN SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
RP TISSUE-Breast carcinoma;
RX Bulwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.
RN [17]
RN CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=97460054; PubMed=9312074;
RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
RT Hantsch F.-G.;
RT "Localization of O-glycosylation sites on glycopeptide fragments from
RT lactation-associated MUC1. All putative sites within the tandem
RT repeat are glycosylation targets in vivo.";
RL J. Biol. Chem. 272:24780-24793(1997).
RN [18]
RN CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RX MEDLINE=99303572; PubMed=10373415;
RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
RT Hantsch F.-G.;
RT "High density O-glycosylation on tandem repeat peptide from secretory
RT MUC1 of T47D breast cancer cells.";
RL J. Biol. Chem. 274:18165-18172(1999).
RN [19]
RN POLYMORPHISM WITHIN THE REPEAT.
RX MEDLINE=21359366; PubMed=11350974;
RA Engelmann K., Baldu S.B., Hantsch F.-G.;
RT "Identification and topology of variant sequences within individual
RT repeat domains of the human epithelial tumor mucin MUC1.";
RL J. Biol. Chem. 276:27764-27769(2001).
RN [20]
RN CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
RX MEDLINE=99211485; PubMed=10197628;
RA Baruch A., Hartmann M.-L., Yoeli M., Aderech Y., Greenstein S.,
RT Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
RA Wreschner D.H.;
RT "The breast cancer-associated MUC1 gene generates both a receptor and
RT its cognate binding protein.";
RL Cancer Res. 59:1552-1561(1999).
RN [21]
RN PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RX MEDLINE=21240104; PubMed=11341784;
RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
RT Harts A.;
RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
RN [22]
RN CHARACTERIZATION.
RX MEDLINE=21836452; PubMed=11847293;
RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
RT Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
RA Stacey M., Lin H.-H., Gordon S.;
RT "Generation of ligand-receptor alliances by 'SBA' module-mediated
RT cleavage of membrane-associated mucin proteins.";
RL Protein Sci. 11:698-706(2002).
RN [23]
RN PHOSPHORYLATION.
RX MEDLINE=95080414; PubMed=7988707;
RA Zrlan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
RT Cytokine receptor-like molecules.";
RL FEBS Lett. 356:130-136(1994).
CC -I- FUNCTION: May play a role in adhesive functions and in cell-cell
interactions, metastasis and signaling. May provide a protective

Query Match

8.7%; Score 651; DB 1; Length 1255;

Best Local Similarity 27.9%; Pred. No. 3.9e-18;
Matches 294; Conservative 74; Mismatches 466; Indels 216; Gaps 50;

QY 274 TSLVNNKETTVEKETTNNKOTDGEKETSIAKETSIAKSD-LAPTSKYLAAPT 332
Db 16 TLTIV-----VTSGHASTPGGEKETSATQSRSSVSPSTKSNVSMTSVLSHS 65
QY 333 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTKEPTPTI---K 372
Db 66 PGSSGSTTQGDVTLAPATPEPSSAATWGDVTSVPTRPALGSTTTPHADVTSAPDNK 125
QY 373 SAPTPKEPAPTTTSAFTTPEKAPTTTKEP-----APTTPEKAPTTTKEPAPTTK 426
Db 126 PAPGSTAPPAHGVTSAPDT--RAPGSTAPPAHGVTSAPDT--RAPGSTAPPAHGVTS 179
QY 427 SAPTPKEPAPTTK---KPAPTTPEKAPPTTKEPPTTKEP--PAPTTKEP----- 474
Db 180 SAPTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTS 239
QY 475 -APTPPEKAPAPTAPE---KPAPTTPEKAPAPTTK---EPAPTTTKEPSPPTTPEKAPPT 525
Db 240 SAPDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTS 299
QY 526 T---TKSAPTTTKEPAPTTTSAFTTPEKAPPTTKEP-----APTTPEKAPAPTTK-- 574
Db 300 SAPDTPRAPAGSTAPPAHGVTSAPDT--RAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAH 356
QY 575 --KPAPTTPEKAPAPTTK---EPAPTTTKEPAPTAPEKAPAPTTTKEPAPTTKLTPTT 628
Db 357 GVTSAFDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAH 416
QY 629 PEKAPPTPEKAPAPTTPELAPTTPEEPTPTTPEEAPATTPKA-----AAPPTPEKAPAPT 684
Db 417 GVTSAFDTPRAPAGSTAP---PAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTA 472
QY 685 PK-----EPAPTTPEKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT-----KKPAP 730
Db 473 PPAHGVTSAPDTPRAPAGSTAP---PAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAP 528
QY 731 KELAPTTTKEPTSTTSKAPAP--TTPKGTAPTTPEKAPAPTTTKEPAPTTKPG---TAPT 784
Db 529 GSTAP--PAHGVTSAPDTPRAPAGSTAPPAHGVTS-----APDTPRAPAGSTAPPAHGVTSAPD 583
QY 785 TLKPEAPAPTT-----KKPAPPELAPTTTKEPTSTTSKAPAP--TTPK---ETAP 828
Db 584 TTPAPGSTAPPAHGVTSAPDTPRAPAGSTAP--PAHGVTSAPDTPRAPAGSTAPPAHGVTSAP 642
QY 829 TTPKEPAPTT-----KKPAPTTPEPTTSEVSTPTTKEPTT-----IHK 871
Db 643 DTPRAPAGSTAPPAHGVTSAPDTPRAPAGS--TAPPAHGVTSAPDTPRAPAGSTAPPAHGVTS 700
QY 872 SPDSSTPELSAEPPTKALENSPEKSGVTTTTPA-----TKPEMTTAKDK--TT 920
Db 701 APDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTS 760
QY 921 ERDRTTPEPTT-----TAAPKMTKEPATTTTEKTESKINATTTQVSTTQDTP--F 971
Db 761 APDTPRAPAGSTAPPAHGVTSAP---DTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAH 816
QY 972 KITTLTKTTTLAPKVVTTKKTTTT--EIMNPEETAKEDRATNSKATTPKOKETKAP 1029
Db 817 GVTSAFDTPRAPAGSTAPPAHGVTSAPDTPRAPAGSTAPPAHGVTSAPDTPRAP--GSTAP 874
QY 1030 KPSTTKKPKTMRKPKPTTPTTPTKMTSTMBELNPTSIAMAMOTTTRPNOT--PNSGLV 1088
Db 875 AHGVTSAPDTP---APGSTAPPAHGVTSAPDTPRA-----PGSTAPPAHGV 918
QY 1089 EVNPKSEDAGAGEBETPMMLRPHVMEVETPMDVYLPRVNOGIIINPMLSDEFINONG 1148
Db 919 TSADTPRAPAGSTAPPAH-----GVTSAFDTPRALDSTA---PVAHVNTSASGS 964
QY 1149 KPVVDGLTTLRNGTLVAFRGHYFMMLSPPSPS 1180
Db 1149 KPVVDGLTTLRNGTLVAFRGHYFMMLSPPSPS 1180

Db 965 ASGSASTLVHNGTSARATTTTPASKSTPSPS 996

RESULT 10
UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benham G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT transduction domains.";
RL J. Cell Biol. 132:835-848(1996).
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RN Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Structural component of the muscle M-line. Myofibril
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -1- SIMILARITY: Contains 1 DBJ-homology (DH) domain.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 5 RGS domains.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U33058; AAB00542.1; -
CC EMBL; AF003131; AAB54132.2; -
CC PDB; 1FHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig C2.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001849; PH.
CC InterPro; IPR007850; RGS.
CC InterPro; IPR000219; RhogEF.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00044; fn3; 1.
CC Pfam; PF00047; Ig; 47.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF05177; RGS; 5.
CC Pfam; PF00621; RhogEF; 1.
CC Pfam; PF00018; SH3; 1.

DR SMART; SM00408; Igc2; 23.
 DR SMART; SM00325; RhogE; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50010; IG_LIKE; 49.
 DR PROSITE; PS50035; IG_LIKE; 49.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KM Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
 3D-structure.
 FT DOMAIN 63 127 SH3.
 FT DOMAIN 152 330 DH.
 FT DOMAIN 342 498 PH.
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
 FT DOMAIN 748 736 IG-LIKE C2-TYPE 2.
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1272 1315 THR-RICH.
 FT DOMAIN 1375 1475 RCSD 1.
 FT DOMAIN 1479 1585 RCSD 2.
 FT DOMAIN 1597 1695 RCSD 3.
 FT DOMAIN 1700 1799 RCSD 4.
 FT DOMAIN 1800 1860 RCSD 5.
 FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
 FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
 FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
 FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
 FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
 FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
 FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
 FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
 FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
 FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
 FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
 FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
 FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
 FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
 FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
 FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
 FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
 FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
 FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
 FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
 FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
 FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
 FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
 FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
 FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
 FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
 FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
 FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
 FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
 FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
 FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
 FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
 FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
 FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
 FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
 FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
 FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
 FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
 FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
 FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
 FT DOMAIN 6275 6368 FIBROBLASTIN TYPE-III.
 FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
 FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
 FT DISULFID 621 POTENTIAL.
 FT DISULFID 2908 POTENTIAL.
 FT DISULFID 3015 POTENTIAL.
 FT DISULFID 3707 POTENTIAL.
 FT DISULFID 3826 POTENTIAL.

FT DISULFID 5092 5157 POTENTIAL.
 FT DISULFID 5298 5350 POTENTIAL.
 FT DISULFID 5508 5560 POTENTIAL.
 FT DISULFID 5616 5669 POTENTIAL.
 FT DISULFID 5722 5764 POTENTIAL.
 FT DISULFID 5836 5901 POTENTIAL.
 FT DISULFID 5946 5998 POTENTIAL.
 FT DISULFID 6036 6171 POTENTIAL.
 FT DISULFID 6421 6486 POTENTIAL.
 FT CONFILCT 2137 2137 A -> P (IN REF. 1).
 FT CONFILCT 2245 2247 ACA -> PKP (IN REF. 1).
 FT CONFILCT 2258 2258 A -> P (IN REF. 1).
 FT CONFILCT 2284 2284 B -> G (IN REF. 1).
 FT CONFILCT 2297 2297 M -> I (IN REF. 1).
 FT CONFILCT 3531 3531 A -> G (IN REF. 1).
 FT CONFILCT 3531 3531 A -> G (IN REF. 1).
 FT CONFILCT 3864 3864 DAGEY -> RRRRI (IN REF. 1).
 FT CONFILCT 3929 3929 A -> V (IN REF. 1).
 FT CONFILCT 5134 5134 A -> P (IN REF. 1).
 FT CONFILCT 5145 5145 T -> S (IN REF. 1).
 FT CONFILCT 5185 5185 G -> A (IN REF. 1).
 FT CONFILCT 5199 5199 K -> N (IN REF. 1).
 FT CONFILCT 5202 5202 L -> F (IN REF. 1).
 FT CONFILCT 5213 5213 F -> L (IN REF. 1).
 FT CONFILCT 6178 6178 A -> G (IN REF. 1).
 FT CONFILCT 6268 6268 K -> E (IN REF. 1).
 SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;
 Query Match 8.4%; Score 629; DB 1; Length 6632;
 Best Local Similarity 23.4%; Pred. No. 1e-16;
 Matches 273; Conservative 156; Mismatches 467; Indels 270; Gaps 44;
 141 NKKTKKTVIESEIEEHSVSENOR-----SSSSSSSSSSSTIMIKSSKSNANRELQK 196
 1176 NKRKRVDOEGARVDRSDTVDGASILTIDATVYSEVNLITSYV--AENTLGAEEGCA 1233
 197 KKKVNDKNNKPTKPKPPVVDAGSGLDNGDKVTTPTDSTTQHNK-----VSTSP 250
 1234 QLTIRPKESVYVER-----QDLSSSEVQKEINQVKEASP 1269
 251 KITTAKPINPRPSLPPNSDTSKEISLTVNKKETVETKETTNNKQSTDGKEKTSANET 310
 1270 EATTV-----ITWETSLTSTKTTMTSTEVTSVGVGVETKESSESATV 1315
 311 Q-----SIETSAKDLPATSKVLAKPTPKKETTGGALLTPKKEPTTPPKS 357
 1316 VIGGGGGVTEGSSISVSKIEVNSKTDSDQDVAEGPKRVSABEEL--PKREVDSRRK 1373
 358 PASTTP--KEPTPTIKAPPTPKKEPAPPTTKSAPPTPKKEPAPPTTKKEPAPT 415
 1374 KSPSPDCKKESPEKTEKRPASPTKGTGEVKS-----PKE-----KSPASTKKEKSPA 1423
 416 TTKEBAPPTTKSAPPTP--KEBAPTPPKKAPPTPKKEPAPPTPKKEPAPT 473
 1424 AEEVSPKPKESPSPTKKEKSPSPKTKGDBEVKSP-----PK--SPTKKE 1471
 474 PAPTPPKKAPAPAPKAPPTPKKEPAPPTPKKEPAPTTPKESBPTT--PKAPAPTTKSAPT 532
 1472 KSPKPEDEVKSPVKK-----KSPDATNIVEVSEETIETKETMTTMEHTHESESRTS 1525
 533 TTKEBAPPTTKSAPPTP---KEBPTTKKEPAPPTPKKEPAP--TPPKAPPTPKKEAPT 587
 1526 VKKETPEKVDKPSPTKOKSPKSIITEBKSVKVKESPKKEBKEKAPASTKKEKSPA 1585
 588 TPKEBAPPTTKKAPAPTAKEPAP-----TPKKEAP--TPPKLPTTPEKLAAPTTP 637
 1586 KPASPTKKSSENVKSPTKKESPKESVVEBELKSPKESPKADDPKSPPTKKEK-----SP 1641
 638 EKPA-----PTTPELAPTTPEBPTPTTPEBAPT--TPKAAAPNTKKEAPPTPKK--P 688
 1642 EKSATEDVKSPTKKEKSPKVEKEKPTSPTKKESSTPKTDEVKSPTKKEKSPQVKEKP 1701
 689 APTTKEBAPPTPKETAPPTPKGTAPPTLKKEPAPPTPKKAPKELAPTTKEPTSTSDK 748

```
Db 1702 ASPTKEKSPKSVVEVYKSPKSP-----EKAEKPKSPKPK-----KSPKSAABE 1751
Qy 749 PAPTPKKTAPTPPKAPPTPKAPPTPKGATTLKBPAPTPPKKAPKAPKAPLAPTTK 808
Db 1752 VKSPKPKKSPKSAABEKPKSP-----TKKSSSVKN-----ADDEVKSPKPKKSP----- 1798
Qy 809 GPTSTSDKPAPTPKETAPPTPKAPPTPKKAPPTPKPPTSPPTSPPTSPPTSPPTSPPT 868
Db 1799 ---EKVEKSPKSPKPK-----KTPKSAABEKSPKPKKSPKSPKSPKSPKSPKSPK 1844
Qy 869 IHKSPDSTPELSAPPTPKALENSKPKGVPPTKTPAATKPBMTTAAKTERDLRTTP 928
Db 1845 KEKSP-EKPEKPKSPKSPK-----KSPGSPKPKKSPKSPKSPK-----PAPKLTREDLQTVN 1897
Qy 929 ETTAAPKMTKBTATTEKTESKTLATTVQVSTTQ-----DTPPKITTLTKTT 980
Db 1898 KTDLHFEVVEHATECKMFLDGKAITTAQGVASKDQFEFRCSIDTTFGSGTV--SV 1955
Qy 981 LAPKVTTKKTTTTEIMNKPEETAKP----- 1007
Db 1956 VASNAAGSVETKELKVLTPRETCKPPTDKLRMEVTKGDTVMQVIALHSPLYKMYQ 2015
Qy 1008 -----KDRATNSKATTPKPO--KPTKAPKKTSTYKPKTPMVRKPKTTP 1050
Db 2016 NGNLEDEKNGVTTIKNEENKSLIIPNAODSGKITVEASNEVGSSESAQLTVNPPSTTP 2075
Qy 1051 T-----PRKMTSTMPRL-----NPTSR-----IAEAMLTTRN-----QTP 1083
Db 2076 IVDGPKSVTTKETETAPKATISGFPAPVTVMINKIVESSRTITTKEDVYTKIS 2135
Qy 1084 NSKIVE---VNPKSEDAAGETPHMLRPHVPEVTPDMQDLYRVNQG-----I 1133
Db 2136 NAKIEGTGTAVTAQNSAGQSGSKQADLKEPVNVAKPKRSQI--TDXYADESEPLRNMLE 2193
Qy 1134 IINPMLSDETN-ICNGKPVDTLTLR 1158
Db 2194 LDGSPGTGVSWLNGOPLTKSDTVQ 2219

RESULT 11
ID PCLO CHICK STANDARD; PRT; 5120 AA.
AC 28-FEB-2003 (Rel. 41, Created)
AT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB Piccolo protein (Aczonin) (Fragment).
GN PCLO OR ACZ.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99439764; PubMed=10508862;
RA Wang X., Kibechull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilmann M.W.
RT "Aczonin, a 550-kD putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin".
RT J. Cell Biol. 147:151-162(1999).
RL
CC -I- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -I- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).
CC -I- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions (By similarity).
CC -I- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -I- SIMILARITY: Contains 2 C2 domains.
```

```
CC -I- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```

EMBL; Y19187; CAB60725.1; -.
HSSP; P04410; 1A25.
DR GO; GO:0045202; C:synaptic junction; ISS.
DR GO; GO:0005509; F:calcium ion binding activity; ISS.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
DR GO; GO:0005522; F:profilin binding activity; ISS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001565; Synaptotagmin.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00399; SYNAPTOTAGM.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2 DOMAIN 1; 1.
DR PROSITE; PS50004; C2 DOMAIN 2; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;
KW Repeat.
KW NON TER
FT DOMAIN 1 258 357 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-X.
FT ZN FING 368 392 C4-TYPE (POTENTIAL).
FT ZN FING 836 859 C4-TYPE (POTENTIAL).
FT DOMAIN 2324 2343 POLY-PRO.
FT DOMAIN 4414 4493 PDZ.
FT DOMAIN 4627 4726 C2 DOMAIN 1.
FT DOMAIN 5003 5094 C2 DOMAIN 2.
SQ SEQUENCE 5120 AA; 560751 MW; A658D9891B65B412 CRC64;

Query Match 7.7%; Score 578.5; DB 1; Length 5120;
Best Local Similarity 20.8%; Pred. No. 6.6e-15;
Matches 333; Conservative 179; Mismatches 542; Indels 549; Gaps 70;

```
Qy 108 HNPSPSSSKAPPPSGASQT---IKSTKSPKPKKTKKYISEETIEHSVSEN 163
Db 5 HHPKQP--GKPPDGGPGISKSRVTVLKEGRAD---GRSPSISIRSEKSRDPEKD 57
Qy 164 QESSSSSSSSSSSTIMKIX-----SKNSANRELOKKLYKDKKDKRTK--- 209
Db 58 QKPSMMPSEFSEANPLSAVTSVVKPSFPDLISDAAHEAGKQKTKQKEQGPEBQR 117
Qy 210 ---KKPTPK--PVPVDEAGSLDNGDFKVTTPDSTTQHNKVSPTSKITAKPINPR--- 261
Db 118 GLAHNPQGGQPKVVGQGG-----FVAPTEQTE--SSKPVVQGGQPPGPKQGQ 164
Qy 262 ---PSLPNSDTSKETSITVNKETVETKETTTNNKQSTIDGKEKTTSAKQSTIEKTS 318
Db 165 KPGSHHPDGSKAEQ-----VKQPPQPRGPKSQLOQSEPTPKGQOQTS--- 212
Qy 319 KDLPATSKVLAKPFPKATTTKGAALTPKKEPTTPTEKERTSTPKERT----- 367
Db 213 ---GPT-----KLPDQPSAK-----TSSQAPPTKDSLOQSGSVKQPSQPARQGFVK 260
Qy 368 PTTIKSAF--TTPPEAPTTTKSAPT--TPKEAPTTTT---KEBAPTPKPEAPTTTKEP 420
Db 261 PSAQQAAGPKQKQPSSEKPTAQTGTPAKQPPRGKPTPLQQTGVKQVPPAGPT---KP 317
Qy 421 APPTTKSAPTPKPEAPPTP--KKPAPTPKPEAPTPKKEPTTPPKK---PAPTKPEA 475
Db 318 SSQTAGAKSLAQPGTLKPPGQOPGPEKPLQQRKASTTQGVESSTPKKTFCLCTTELL 377
```

```

QY 476 PTTKPK-----PAPPAK-----PAP 492
   ||:|
   ||:|
Db 378 LHTPEKANNVTCTQCHTVVSLGCFNPNPHITELKELCLNCOMORALGSDLASGHCPCP 437
QY 493 TTPKEPAPTTPKEPAPTTPKEPPT---TPKEPAPTTTSAPPTTPKEPAPTTTPKAAPT 548
   ||:|
   ||:|
Db 438 QLP-----PRKQKPTPTASTAKPSBQLPGQCKDASPRKPSQOADSCKVPF---QKQPSM 491
QY 549 PKEPSPPTTPKEPAPTTPKEPAPTTPKPAPTPK---EPAPTTTPKEPAPTTTPKKAAPTA 604
   ||:|
   ||:|
Db 492 PGSP-EVKSQO-----THAEPSTDT---GQOQIDSTPSDQVKEPTQAEKQNPQSIKPTMDIV 544
QY 605 PKEPAPPTTPKEPA-----PTPKLTPTTPKEKLAPTTPKEPAPTTPKEEALTTBEPPTPT 660
   ||:|
   ||:|
Db 545 PTSAAGCVQODLADPOSPSTQCKVT---DSMPETTTKPADTHPADKDS---KLPQV 558
QY 661 PEEBPAPTPKAAAPNTTPKEPAPT-TPKEPAPT---TPK---EPAPTTTPKEPAPTTPK-GT 712
   ||:|
   ||:|
Db 599 SRQ---KSDPKLASGSAKSDAKTQKPEEPAPVDDPKQTKAPAKPDTPKAPKGPQAGT 656
QY 713 APPTLKEPAPTTPKPAKPAKLAPTTTPKEPTS-----TTSDDKPAPTTPK----- 755
   ||:|
   ||:|
Db 657 GPR-----PTSAQAPAPQPOQKTPBEOGRPSRLNKGITDAPKQPTTPPOETVNGKLP 709
QY 756 -----GPAPTTPKEPAPT-----PKEPAPTTTPKGA 782
   ||:|
   ||:|
Db 710 GPCASIFSOASSLISIRAGSGSOTSGPAPATKOPPOPSOPASOAPKEAQAQAPPPKA 769
QY 783 PTTLKEPAPTTPKPAKPAKLAAPT-TTKGP-----TSTSDKPAPTTPKEPA 827
   ||:|
   ||:|
Db 770 APFKKETTPLASEKLPMSDSLTITKSGDLEKKPLAKDSKHQTLMAKKAAPALSQEK 829
QY 828 ----- 827
Db 830 SOPKVSCLPCKTGLNIGSKDPPNNTCTECKKVCNLCGFNPNPHIVEQEMCLNCGTO 889
QY 828 -----PTTPKEPAPTTP-KKPAPTTPETP-----PPT---TSEVST 859
   ||:|
   ||:|
Db 890 RAMSGOLGDMGKVPRLKLGPSOVSKAPRATPOKQVPVANSVHSQKSTPTTPATTPKEE 949
QY 860 PTTTPKEPTTIHKSPPDESTBELSAEPTPKAL-ENSPKEPVPPTTKTPPAATTPKEMTTAKD 917
   ||:|
   ||:|
Db 950 PSVPEKVPRLQOGKLEKT--LSADKIQOGIQKEDASKGS-KLFPKPSADKQVSGOKED 1006
QY 918 -----KTERDLATTPETTTAAPKTKZ----- 940
   ||:|
   ||:|
Db 1007 SRLQOTKLTCTPSSDKILNGVQKEDIKFOEAKLAKIPSAADKILHRLQKEDPKLQCKMAK 1066
QY 941 -----TATTEKTTES-----KITATTT-----Q 959
   ||:|
   ||:|
Db 1067 ALSADKIOPBAQEDVQLOEVRLSKAVSADKIQHGIQKIDLNQHVLIKETSSVEKIQEAO 1126
QY 960 VTSITTTQDTPPKITTLTKTTTLAPKTYTT-KKTIITTEIMNKE-----ETAKPKDRATNS 1014
   ||:|
   ||:|
Db 1127 KSKKIQODKLP--KTLSEDKIPATVSSOHKULSSEBKKLELEKSPHKKDKEQI 1183
QY 1015 KATT-----PKQKPTKAPKKTPTSTKQKTPMPVRVKKKPTTPPRKA 1055
   ||:|
   ||:|
Db 1184 TLETTGHTIIOKAVEAPCDKLEKQOEDKEDLTGIPQWVSKPEKKEEETKTPVP--- 1240
QY 1056 TSTMPELNPTSRIEAM-----LOTTTRPNQTPNSKLEVNANKSDAGABEETHML 1109
   ||:|
   ||:|
Db 1241 VSRLEPSDIVEAVREKIEKEDKSDTSSGQCKSPGLSDTGVSSGJSSSJEIIP--- 1296
QY 1110 RPHVPEVYTPDMVLPVFNQCIILNPMLSDETNI CNKGPVDTGLTTLNGLTLVAFRGHY 1169
   ||:|
   ||:|
Db 1297 -----SHLPSDEKDLPREPSQ---KDLISGE----- 1319
QY 1170 FFMLSFPSPSPARATTEWAGISPIDYPTTRCNCGKTPFFKDS-----QYRFPNDIK 1224
   ||:|
   ||:|
Db 1320 -----SPSPSPS-----DLAKKESYLSILEAQASTLTDEKSVKREKELYETVSEQTK 1365

```

```

QY 1225 DAGYKPIFKGGLTQGLVALSTAKYKNWPEVYFFK-----RGGSIQOYIK 1274
   ||:|
   ||:|
Db 1366 DQHKTYEL-----EYTPESYSSDEEDLAIQEGERTTIAADSKGASSTQDYK 1412
QY 1275 QEB-VQKCPGR--PALNYPVYGEMTOYRRRRFERAIGPQOT 1313
   ||:|
   ||:|
Db 1413 EEDGNDTTPARQRYDSVEDSSSENSPVPRRRRRASVSSSS 1455

RESULT 12
ZAN HUMAN
ID ZAN HUMAN STANDARD; PRT; 2812 AA.
AC Q9Y493; 090218; 096L85; 096L86; 096L88; 096L89; 096L90;
AC Q9BXN9; 098283; 09B284; 09B285; 09B286; 09B287; 09B288;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC TISSUE=Testis;
RA Cheung T.L., Masiar M.J., Cornwall G.A., Hardy D.M.;
RT "Multiple intra-species variants of human zonadhesin."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RC MEDLINE=99018118; PubMed=9799793;
RA Glockner G., Scherer S., Schatevov R., Boright A.P., Weber J.,
RA Tsui L.-C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUGT1 loci
RT reveals 17 genes."
RL Genome Res. 8:1060-1073(1998).
RN [3]
RP SEQUENCE OF 1810-2812 FROM N.A. (ISOFORM 1).
RC MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,
RA Cheung T.L., Hardy D.W., Schwartz S., Scherer S.W., Tsui L.-C.,
RA Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHE/TFE2 region on human
RT chromosome 7q22 with the orthologous region on mouse chromosome 5."
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [4]
RP SEQUENCE OF 2375-2683 FROM N.A. (ISOFORM 7).
RC TISSUE=Testis;
RA MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Hartum T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN)."
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Name=3;
CC IsoId=Q9Y493-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9Y493-2; Sequence=VSP_001430, VSP_001431;
CC Name=2;
CC IsoId=Q9Y493-3; Sequence=VSP_001428, VSP_001429;
CC Name=4;
CC IsoId=Q9Y493-4; Sequence=VSP_001424, VSP_001425;
CC Name=5;
CC IsoId=Q9Y493-5; Sequence=VSP_001420, VSP_001421;

```



```

QY 317 SAKDLAPTSKVLAKPTPKAETTKGPAALTTPKEPTTPPKPASTTPKEPT-----PTTI 371
DB 578 SVTTEKPT-----VPEKEPTIPTEKPTISTEK---PTIPSEKNMSEKTIPTSEKPTII 629
QY 372 KSAPPTPKBPAPTTTSKAPTPPKBPAPTTTPKEBPAPTTTPKEBPAPTTTPKSAPT 430
DB 630 TEKPTIPSE-KPTIPSEKPTISTEKPTVPTPE--PTTPEETTTSMEBPVPTPEKSIPT 666
QY 431 TPKEBPAPTTTPKEBPAPTTTPKEPTTPPKBPAPTTTPKEBPAPTTTPKEBPAPTKP 490
DB 687 --EKPSIPTEK---PTISMEETIISTEKPTISPEKPTIPTKE--PTIPSEKSTISPEK- 737
QY 491 APTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKSAPTTPKSAPTTPK 550
DB 738 -PTTPE-KPTIPTE-KPTIPSE-KPTIPTEKPTISTEKPTIPTKEKPTIPTKE--PTIPT 791
QY 551 EBPPTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKPAPTAP 605
DB 792 EKPTISTEE--PTTPEETTISTEKPSIPEKPTLPTPEETTTSVEETTISTEKLTIPTM-- 847
QY 606 KEBPAPTTPKETATP-----TPKULTPTTPPKLA-----PTTPEKBPAPTTPEELAPTTPE 655
DB 848 EKPTISTEKPTIPTKEPTISTEKPTIPT- EKLTIPTEKPTIPTPEETTISTEKL--TIPTE 904
QY 656 PPTTPEBPAPTTPKAAPNTPKBPAPTTTPKEBPAPTTTPKEBPAPTTTPKGTAPT 715
DB 905 KPTIPSEKPTISTEK-----PTTPE-KPTIPTE-----ETIISTEKLTIPT 945
QY 716 TKBPAPTTPKKAPKELAPTTTPKEPTSTT-----SDKBPAPTTKGTAPTTTPKEBPAPT 769
DB 946 --EKPTIPSEKPTIPTKEPTISTEKPTIPTKEKPTIPTKEKPTIPT- EKLTAPR 1001
QY 770 PKBPAPTTKGTAPTTTPKEBPAPTTTPKAPKELAPTTTPKGTSTSDKBPAPTTPKGTAPT 829
DB 1002 PPHPSPTA-TGLALVMSPHAPSPMTSV---ILGTTTSSSTSEKCPNARYESCACPA 1057
QY 830 TPKEBPAPT 837
DB 1058 SCKSPAP 1065

RESULT 13
PPI_MYTBD
ID_PPI_MYTBD STANDARD; PRT; 875 AA.
AC Q23460;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adhesive plaque matrix protein (Polypheonolic adhesive protein) (Foot
protein 1) (MERP1) (Fragment).
GN PPI.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidae; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91025829; PubMed=1367451;
RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
RT "Structural and functional repetition in a marine mussel adhesive
protein.";
RL Biotechnol. Prog. 6:171-177(1990).
RN [2]
RP PARTIAL SEQUENCE AND POST-TRANSLATIONAL MODIFICATIONS.
RX MEDLINE=83135732; PubMed=6298211;
RA Waite J.H.;
RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
hydroxyproline-containing decapeptide in the adhesive protein of the
mussel, Mytilus edulis L.";
RL J. Biol. Chem. 258:2911-2915(1983).
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S

```

```

CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY
CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC (DOPA) DERIVED FROM TYROSINE.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
CC EMBL; X54422; CAA38294.1; -.
CC PIR; S23760; S23760.
CC InterPro; IPR002964; Adhesive_plaq.
CC InterPro; IPR006031; XYPEX.
CC Pfam; PF02162; XYPEX; 55.
CC PRINTS; PR01216; ADHESIVEI.
CC Repeat; Hydroxylation.
CC KM NON TER 1
CC FT 67 870 TANDEN REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
CC FT DOMAIN P-P-[ST].
CC SQ SEQUENCE 875 AA; 100412 MW; 6EA85312748CACE CRC64;

Query Match 7.4%; Score 559; DB 1; Length 875;
Best Local Similarity 28.7%; Pred. No. 8.4e-15;
Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;

QY 241 TQHNKVSISPKITAKPINP-----RP--SLPNSDTSKETSILTVNKEETVET 286
DB 1 TKEBPYKVKVTSYSAKPYKPTTYPQPLKKKVDRPTKSYPTTG-SKTNVLLPLAKKLSYK 59
QY 287 KETTTNKOTSDG--KEKTT--SAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTKG 341
DB 60 PIKTTYNAKNTNPPYVKPKMTYPTPKPSYPTPKSKPTPKYKPTTYPAKPSY-- 117
QY 342 PALTPKEPTTPTEK-----PASTTPKEPTPTTISAPTT-----KEBPAPTTTSAPT 392
DB 118 PSYKPKTYPTTYPKLTYPTTYKPSYPTTYKPSYPTTYKPSYKTYKPSYKAKPSY 177
QY 393 PKBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKSAPTTPK--BPAPT 439
DB 178 P-----PTYKAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPTTYKAKPSYPTTYKAKPSY 233
QY 440 PKKPA--PTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAP 488
DB 234 KAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTY 293
QY 489 KPA--PTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAPTT 538
DB 294 KPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTY 353
QY 539 PTTTSAPTTTPKPSPTTPKEBPAPTTTPKEBPAPTTTPKPA--PTTPKEBPAPTTTPKEBP 594
DB 354 PTTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSY 409
QY 595 TTTTKBPAPTPAKBPAPTTTPKETAPTTPPKULTTTPBEKLAPTTPBEKAPPTTPEELAP 654
DB 410 PPTTYKAPTYKAKP-----TYPST-YKAKPSYPTTYKAKPSYPTTYKAKPSYPTTY 654
QY 655 BPTTTPKEBPAPTTPKAAPNTPKBPAPTTTPKEBPAPTTTPKEBPAPTTTPKEBPAPT 708
DB 458 KPT-----YSTTKA-----KSYPAKYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTY 708
QY 709 PKGTAPTTTPKEBPAPTTTPKAPKELAPTTTPKEPTSTSDKBPAPTTTPKGTAPT--PK 764
DB 507 PLUTYKPTVK--PKPSYPSYKKTTPYPTVK-----PKSYPTTYKAKPSY 551

```


RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosvelde F.,
 RA Moshynski W.;
 RT "Sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit.";
 RL Gene 68:307-314(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Shneidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT revealed by cDNA and genomic sequences.";
 RL Brain Res. 464:217-221(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RA Carden M.J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: I, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, M24496; AAA39813.1; -;
 DR EMBL, M23149; AAA39813.1; JOINED.
 DR EMBL, M24494; AAA39813.1; JOINED.
 DR EMBL, M24495; AAA39813.1; JOINED.
 DR EMBL, M35131; AAA39809.1; ALT_FRAME.
 DR EMBL, Z31012; CAA83228.1; -;
 DR PIR, J03668; QPMH.
 DR MGI, MGI:97309; Negh.
 DR InterPro, IPR001664; IF.
 DR Pfam, PF00038; filament; 1.
 DR PROSITE, PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 Repeat.
 FT DOMAIN 1 97 HEAD.
 FT DOMAIN 98 408 ROD.
 FT DOMAIN 409 1087 TAIL.
 FT DOMAIN 436 517 GLU-RICH (ACIDIC).
 FT DOMAIN 519 865 50 X 6 AA TANDDEM REPEATS OF K-S-P-A-E-A.
 FT DOMAIN 887 1087 GLU/LYS-RICH.
 FT DOMAIN 98 129 COIL 1A.
 FT DOMAIN 130 141 COIL 1B.
 FT DOMAIN 142 239 COIL 1B.
 FT DOMAIN 240 261 LINKER 12.
 FT DOMAIN 262 283 COIL 2A.
 FT DOMAIN 284 287 LINKER 2.
 FT DOMAIN 288 408 COIL 2B.
 FT DOMAIN 409 488 COIL 2B.
 FT CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).
 FT CONFLICT 199 199 A -> AR (IN REF. 2 AND 3).

FT CONFLICT 281 281 S -> T (IN REF. 2 AND 3).
 FT CONFLICT 492 492 L -> G (IN REF. 2 AND 3).
 FT CONFLICT 551 551 P -> PREAKSP (IN REF. 3).
 FT CONFLICT 689 712 MISSING (IN REF. 3).
 FT CONFLICT 714 714 G -> A (IN REF. 3).
 FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).
 FT CONFLICT 843 843 T -> N (IN REF. 2 AND 3).
 SQ SEQUENCE 1087 AA; 116612 MW; 57BAC76A38ED1C89 CRC64;
 Query Match 7.3%; Score 551; DB 1; Length 1087;
 Best Local Similarity 26.8%; Pred. No. 2e-14;
 Matches 224; Conservative 96; Mismatches 358; Indels 186; Gaps 43;
 143 KTKKVIIESE--EITEHVS--ENESSSSSSSSSSSTIWKIKSKNSAARELOKTK 199
 331 KSTKSLERQSELEDRHOADIASYODAIQOLDSLRNWKWM-----AAQUREYODLIN 385
 200 VK-----DNKNRTKKKPPKPPVNVAGSGLDNGDFKVTTPSTTQHNK 245
 386 VQWALDIEIAARKLLEGECECHIGFSP-----FSLTEGLPKI--PSIST--HIK 432
 246 VSTSPKITTAKPINDPRLPNSDTSKETSITVKNKETTVEKTTTNNKQSTDGKETT 305
 433 VKSEEMIKVVE-----KSEKETVIVEGQTEIRVTEGVTEEDKBAQCEGHE 480
 306 SAKETOSTIKTSAKOLAPTSKVLAKPTKRAETTGPAITTKETPTTPKREPASTTPKE 365
 481 AEGERKEEKEELAAITSPAEBAASPEKETSIVVEAKSPGEAKSPGEAKSPA---BA 536
 366 PPTTIKSNAPTPKPEAPPTTKSAPTTPKEP--APTTKEPAPTPKPEAPTTKEP--- 420
 537 KSPGSAKS--PGSAKSPGEAKSPAEPKSPAEPKSPAEPKSPA--EKSRA--TVKSPGEAK 591
 421 APTTKSAPTTPKEPA---PTTPKPA---PTTPKPA---BTTPKEP---TPTTP 464
 592 SPSEAKS--PAEAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 650
 465 KEPAATTEBPATTPKEPA---PTAPKP---APTTKEPAPTPKPEAPTTKEPSP 516
 651 KSPA--EKSSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA--KSPA---EVSP 701
 517 TTPKEPA---PTTKSAPTTPKEPAPTTKSAPTTPKE--PSPTTKKEPA---PTTPKE 567
 702 AEAKSPAEEKSPAEEKS--PAVKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA 760
 568 PAPTPPKAPPTTPKEPAPTTPKEPAPTTKKAPPAEPKPEAPTTPKETAPTPKKTPT 627
 761 PIEVSPKPAKTPVVEGAKSPAEEKSPAEEKSPAEEKSPAEEKSPAEEKSPA--KSPVGEAK 819
 628 TPKEKAPPTTPKEPAPTTPKEPAPTTPKEPTTPKE--BAPTTKKAAPPTTPKEPAPT-- 684
 820 PPEKAPPAIVKSPAEQIVQEBATVDIRP--PPOVKSPAEKKAESP--KEEAKTSK 875
 685 --PKEPAPTPP-----KEPAPTPKETAPTTPPGTAPTTKBAPTPPKKPAKELAP 735
 876 VAPKEEVKSPVKEVAKKEPKKYBEKTLPTPTKEAKESKDAAPKAPKVEBKKE 935
 936 TPTKEPKDSTAEEKKEAGEKKKAVASBEETPAKGVGEAA--KKEKETETTKTAEDTK 993
 736 TTKKEPTSTSD--KPAPTPKGAAPTPPKPEAPTPPKPEAPTPPKGAAPTTLKEPAPT 793
 994 AKEPS-----KPTETKP-----KKEEPPAPPEK----- 1017
 854 TSEVSTPTTKETPTTHKSPDESTPELSAEPKALENSPEKPGVPTTKPATKPEMTT 913
 1018 -----KOTKEKTESKKEPK-----PK-MAEKVVEDDLSKSP--SKPTKEK 1059
 914 TAKDKTERDLKTTPETTTAAAPKMTKETATTTK 947
 1060 AEKSSSTQKESQPE-----KTTEDKATGEEK 1087

Mon Dec 8 09:50:45 2003

us-09-556-246-1.rsp

Page 22

Search completed: December 8, 2003, 09:35:50
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 8, 2003, 09:19:41 ; Search time 54 Seconds
(without alignments)
4126.891 Million cell updates/sec

Title: US-09-556-246-1

Sequence: 1 MAMKTLPIVLLLSVFIQ.....ARAITRSGQLSKWVNCNP 1404

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

```
A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7526	100.0	1404	13	MSF precursor. Sy
2	7526	100.0	1404	22	Human megakaryocyt
3	7526	100.0	1404	22	Human megakaryocyt
4	7523	100.0	1415	22	Novel human secret
5	6950	92.3	1299	22	Human EST encoded
6	3484	46.3	902	22	Human MSF-derived
7	2929	38.9	551	22	Human testes-deriv
8	2920	38.8	546	22	Human testes-deriv
9	2850.5	37.9	538	23	Human cartilage super

10	2757	36.6	513	22	ABU53254	Human testes-deriv
11	2197	29.2	452	16	AAR80041	Human megakaryocyt
12	1707.5	22.7	472	22	AAB60569	Bovine MSF ortholo
13	1545	20.5	292	23	AAU11261	Human HARO polypep
14	1188	15.8	5179	22	AAK24516	C9orf9 predicted am
15	1188	15.8	5179	24	ABP53265	Human colon tumour
16	1097	14.6	214	22	ABU53255	Human testes-deriv
17	1012.5	13.5	717	22	ABU51444	Human testes-deriv
18	981	13.0	763	21	AAQ38942	Arabidopsis thalia
19	968.5	12.9	188	23	AAO18833	5' cartilage super
20	950	12.6	1664	19	AAK41006	C. thermocellum OI
21	946.5	12.6	1049	22	ABB61364	Drosophila melanog
22	927.5	12.3	778	22	ABU53143	Human testes-deriv
23	927	12.3	770	22	ABU53141	Human testes-deriv
24	902	12.0	1795	22	ABB69806	Drosophila melanog
25	900.5	12.0	717	22	ABU53145	Human testes-deriv
26	900.5	12.0	717	22	ABU53146	Human testes-deriv
27	900.5	12.0	717	22	ABU53147	Human testes-deriv
28	900.5	12.0	717	22	ABU53148	Human testes-deriv
29	900.5	12.0	717	22	ABU53149	Human testes-deriv
30	900.5	12.0	717	22	ABU53150	Human testes-deriv
31	900.5	12.0	717	22	ABU53151	Human testes-deriv
32	890	11.8	745	22	ABU53142	Human testes-deriv
33	889	11.8	692	22	ABU53155	Human testes-deriv
34	848.5	11.3	745	22	ABU53154	Human testes-deriv
35	818.5	10.9	695	22	ABU53152	Human testes-deriv
36	805	10.7	1325	22	ABG49735	Human liver peptid
37	805	10.7	1325	22	ABB29725	Peptide #2376 enco
38	805	10.7	1325	22	ABB34897	Peptide #2403 enco
39	805	10.7	1325	22	ABB20314	Protein #2313 enco
40	805	10.7	1325	22	AAK55707	Human brain expres
41	805	10.7	1325	22	AAK68085	Human bone marrow
42	805	10.7	1325	22	AAK03645	Peptide #2327 enco
43	805	10.7	1325	23	ABG37612	Human peptide enco
44	782.5	10.4	2112	22	ABB60403	Drosophila melanog
45	776	10.3	395	22	ABU53160	Human testes-deriv

ALIGNMENTS

RESULT 1					
AAR26049					
ID	AAR26049 standard; Protein; 1404 AA.				
XX					
AC	AAR26049;				
XX					
DT	25-MAR-2003 (updated)				
DT	02-FEB-1993 (first entry)				
XX					
DE	MSF precursor.				
XX					
KW	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;				
KW	stability; proteolytic cleavage; adhesion; alternative splicing.				
OS	Synthetic.				
XX					
FH	Key				
FT	Region				
FT	Location/Qualifiers				
FT	1..26				
FT	/label= Exon_I				
FT	26..67				
FT	/label= Exon_II				
FT	67..107				
FT	/label= Exon_III				
FT	107..157				
FT	/label= Exon_IV				
FT	157..200				
FT	/label= Exon_V				
FT	200..1141				
FT	/label= Exon_VI				
FT	1411..1166				
FT	/label= Exon_VII				
FT	1166..1212				
FT	Region				

THIS PAGE BLANK (USPTO)

FT	Region	/label= Exon_VIII 1213..1266
FT	Region	/label= Exon_IX 1266..1331
FT	Region	/label= Exon_X 1331..1373
FT	Region	/label= Exon_XI 1373..1404
FT	Region	/label= Exon_XII
XX		
PN		
XX		
PD		
XX		
PP		
XX		
PR		
X		
A	(GENY) GENETICS INST INC.	
TX		
TX	Clark SC, Geener TG, Hewick RM, Jacobs K, Turner K,	
OR	MPI, 1992-284660/34.	
N-PDB;	AAQ27223.	
XX		
FT	New human mega-karyocyte stimulating factors - for treating	
PT	immune deficiencies, cancer, exposure to radiation or drugs,	
FT	bacterial and viral infections, etc.	
XX		
P8	Claim 1, 2 and 3; Fig 1; 87pp; English.	
XX		
CC	The sequence given is a full length translation from the megakaryocyte	
CC	stimulating factor (MSF) precursor. The sequence covered by exons II,	
CC	III and IV encodes megakaryocyte stimulating factor (MSF). This	
CC	sequence is modified by the addition of an N-terminal sequence encoding	
CC	a secretory leader, an initiating methionine preceeding exon II and a	
CC	terminating codon following exon IV. The cDNA sequence given contains	
CC	sequences derived from human megakaryocyte colony stimulating factor	
CC	(meg-CSF). Exon I contains the initiating methionine, and encodes a	
CC	classical mammalian protein secretion signal sequence. The sequence	
CC	encoding the original meg-CSF includes exons II-IV and is thought to	
CC	terminate in the region between amino acid residues 134 - 147. The	
CC	primary transcript of this gene may be cleaved in different ways to	
CC	yield a family of mRNA's each encoding a cleaved MSF protein. Exons	
CC	V and VI are thought to be related to the activity of the factor and	
CC	are also implicated in the stability, folding and processing of the	
CC	molecule. These exons are also thought to play a role in the observed	
CC	synergy of MSF with other cytokines. Exons V - XII are believed to be	
CC	implicated in the processing or folding of the appropriate structure of	
CC	the resulting factor. ie. one or more of these exons may contain	
CC	sequences which direct proteolytic cleavage, adhesion, organisation of	
CC	the cellular matrix or extracellular matrix processing. Both naturally	
CC	occurring and non-naturally occurring MSF's may be characterised by	
CC	various combinations of alternatively spliced exons from this sequence,	
CC	with the exons spliced together in differing orders to form different	
CC	members of the MSF family.	
CC	(Updated on 25-MAR-2003 to correct PN field.)	
XX		
SQ	Sequence 1404 AA;	
	Query Match 100.0%; Score 7526; DB 13; Length 1404;	
	Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0	
	Matches 1404; Conservative 0;	
OY	1 MAMKTLPIYLILLLISVFIQQVSSODLSGACAGCGEGSRDATCNCDYNCOHYMECCPDF 60	
DB	1 MAMKTLPIYLILLLISVFVIQQVSSODLSGACAGCGEGSRDATCNCDYNCOHYMECCPDF 60	
OY	61 KRVCFAEIASGCKRCPSFERGRBCDDAACOKKYDKCCPYESFCAEVAHNPSPSSKKAP 120	
DB	61 KRVCFAEIASGCKRCPSFERGRBCDDAACOKKYDKCCPYESFCAEVAHNPSPSSKKAP 120	
DB	61 KRVCFAEIASGCKRCPSFERGRBCDDAACOKKYDKCCPYESFCAEVAHNPSPSSKKAP 120	

Qy	121	PSGASQTIISTTKRSPKPNKKTKKVVISEETIEHSHSENOESSSSSSSSSTIIN	180
Db	121	PSGASQTIISTTKRSPKPNKKTKKVVISEETIEHSHSENOESSSSSSSSSTIIN	180
Qy	181	KIKSSKNSAANREIÖKKLVKDNKKKRTKKKPTPKPEVVDIAGSGLDNGDFKYYT	240
Db	181	KIKSSKNSAANREIÖKKLVKDNKKKRTKKKPTPKPEVVDIAGSGLDNGDFKYYT	240
Qy	241	TOHNKYSTSKITTAAPINRPSLPPNSDTSKSTSLVNMKEITVEKETTNNKQTS	300
Db	241	TOHNKYSTSKITTAAPINRPSLPPNSDTSKSTSLVNMKEITVEKETTNNKQTS	300
Qy	301	KEKTSIAKEIOSIIEKTSAKDLAFTSKVLAPTEKAETTTGPAALTTPKEEPTTPK	360
Db	301	KEKTSIAKEIOSIIEKTSAKDLAFTSKVLAPTEKAETTTGPAALTTPKEEPTTPK	360
Qy	361	TTPEKEPTPTTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPT	420
Db	361	TTPEKEPTPTTIKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPT	420
Qy	421	APTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK	480
Db	421	APTTPKSAPTTPKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK	480
Qy	481	EPAPTAPEKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT	540
Db	481	EPAPTAPEKKAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPT	540
Qy	541	TTASAPTTPEBAPTTPKKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK	600
Db	541	TTASAPTTPEBAPTTPKKEBAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK	600
Qy	601	APTAPKEBAPTTPKETAAPTTPKULPTTPPEKLAPTTPKEBAPTTPBEELAPTT	660
Db	601	APTAPKEBAPTTPKETAAPTTPKULPTTPPEKLAPTTPKEBAPTTPBEELAPTT	660
Qy	661	PEBAPPTTPQAAANPTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK	720
Db	661	PEBAPPTTPQAAANPTTPKEBAPTTPKKEBAPTTPKKEBAPTTPKKEBAPTTPK	720
Qy	721	APTTPKKAPEKELAPTTTKEPTSTTSDBKAPPTPKGAAPTTPKEBAPTTPK	780
Db	721	APTTPKKAPEKELAPTTTKEPTSTTSDBKAPPTPKGAAPTTPKEBAPTTPK	780
Qy	781	TAPTTLKEBAPTTPKKAPEKELAPTTTGGPTSTTSDBKAPPTPKETAAPTTPK	840
Db	781	TAPTTLKEBAPTTPKKAPEKELAPTTTGGPTSTTSDBKAPPTPKETAAPTTPK	840
Qy	841	KPAPTTPPEBAPTTPSEVSTPTTKEPTTIHKSPESHPESLSABTPKALNSPK	900
Db	841	KPAPTTPPEBAPTTPSEVSTPTTKEPTTIHKSPESHPESLSABTPKALNSPK	900
Qy	901	TKTPAARKEPMTTAAKOTTERDLRTTPETTTAARPKMTKEATTEKTTESKITA	960
Db	901	TKTPAARKEPMTTAAKOTTERDLRTTPETTTAARPKMTKEATTEKTTESKITA	960
Qy	961	TSTTTTQOTTPPKITTLKTTTLAPRVTTTKTITTEILMNPBEBAKRDRAVNSK	1020
Db	961	TSTTTTQOTTPPKITTLKTTTLAPRVTTTKTITTEILMNPBEBAKRDRAVNSK	1020
Qy	1021	POKPTTKAPKPESTSKKEKTPMRYVAKPTTETTPRKMTSTWELNPTSTR	1080
Db	1021	POKPTTKAPKPESTSKKEKTPMRYVAKPTTETTPRKMTSTWELNPTSTR	1080
Qy	1081	QTPNSKLVBNPKSEDAAGABGETPHMLRPHVFMPEVTPDMDYLPVYPNOGII	1140
Db	1081	QTPNSKLVBNPKSEDAAGABGETPHMLRPHVFMPEVTPDMDYLPVYPNOGII	1140
Qy	1141	DETNIQNGKAPDQGLTTLRNGLVLVAFRGHYFVMLSPSPSPSARITIEWG	1200
Db	1141	DETNIQNGKAPDQGLTTLRNGLVLVAFRGHYFVMLSPSPSPSARITIEWG	1200
Qy	1201	RCNEGKTEFFPKDSQYWRFTNDIKDAGYKPIEFKFGGLTGQIIVALSTAKYK	1260

THIS PAGE BLANK (USPTO)

DB 1201 RCGCBGTFFPKDSQVWRFTNDIKDAGYKPIFFGFGGLTGQVLAALSTAKYKMWESVY 1260
QY 1261 PFRGGSIOQYIYKQBPVQCPGRRPALNTPYVGMTQVRRRFRERAGISQHTTRIQY 1320
D 1261 PFRGGSIOQYIYKQBPVQCPGRRPALNTPYVGMTQVRRRFRERAGISQHTTRIQY 1320
QY 1321 SPARLAVODGVLHNEKVSILMRGLPNVVTSAISLPIRKKPDGYDYAFSPKQYVNDV 1380
DB 1321 SPARLAVODGVLHNEKVSILMRGLPNVVTSAISLPIRKKPDGYDYAFSPKQYVNDV 1380
QY 1381 PSRTARALITRSQGTLSKWNVNC 1404
DB 1381 PSRTARALITRSQGTLSKWNVNC 1404
RESULT 2
AAB60568
ID AAB60568 standard; Protein, 1404 AA.
AC AAB60568;
XX
DT 27-APR-2001 (first entry)
XX
DB Human megakaryocyte stimulating factor (MSF, CACP).
XX
KW Human, CACP protein, campodactylly-arthropathy-coxa vara-pericarditis;
KW MSF, megakaryocyte stimulating factor; synovial lubricant;
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
KW anticharctic.
XX
OS Homo sapiens.
XX
PN MO200107068-A1.
XX
PD 01-FEB-2001.
XX
PE 21-JUL-2000; 2000MO-US20002.
XX
PR 23-JUL-1999; 9905-0145328.
PR 19-JUL-2000; 200005-0145328.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Matman ML;
XX
DR WPI; 2001-182721/18.
XX
PT New composition comprising the campodactylly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints -
XX
PS Example 1; Page -; 34pp; English.
XX
CC The invention relates to a method of treating osteoarthritis via the
CC administration of a composition comprising the campodactylly-arthropathy-
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC The composition may further comprise a local anesthetic. The composition
CC of the invention may be administered via intra-articular or intravenous
CC injection. The human CACP protein is identified in the invention as
CC being megakaryocyte stimulating factor (MSF). The gene encoding
CC "CACP" protein (MSF) is located on chromosome 1q25-31, and mutations in
CC this gene are responsible for the heritable disorder campodactylly-
CC arthropathy-coxa vara-pericarditis, in which patients have synovial
CC hyperplasia without evidence of inflammation. CACP protein (MSF)
CC acts as a synovium lubricant, and can be used to lubricate tissue and
CC joints in the treatment of osteoarthritis. The composition may be
CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
CC loss of range of movement or joint damage). The present sequence
CC represents human megakaryocyte stimulating factor (MSF, CACP protein).
CC Note: This sequence is not given in its entirety in figure 4 of the
CC specification, although a Genbank accession number was given. This
CC sequence was therefore obtained from Genbank (U70316).

XX	Sequence	1404 AA;	SQ
QY	Query Match	100.0%; Score 7526; DB 22; Length 1404;	
DB	Best Local Similarity	100.0%; Pred. No. 0;	
DB	Matches 1404; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MAMKTLPIYLILLISVFVLIQVVSQDLSCAGRCCEGYSRDATCNCVNCQHYMECCPDF 60		
DB	1 MAMKTLPIYLILLISVFVLIQVVSQDLSCAGRCCEGYSRDATCNCVNCQHYMECCPDF 60		
QY	61 KAVCTAELSCGRCESEFERGECDDAQCCKYDKCCPYESFCABVHNPISPPSKAP 120		
DB	61 KAVCTAELSCGRCESEFERGECDDAQCCKYDKCCPYESFCABVHNPISPPSKAP 120		
QY	121 PPSGASQTIKSTTKSPKPNKKTKVLESEITEHSHSVSNOSSSSSSSSSSTIM 180		
DB	121 PPSGASQTIKSTTKSPKPNKKTKVLESEITEHSHSVSNOSSSSSSSSSSTIM 180		
QY	181 KIKSSKNSAANRELQKLVKONKKNRTKKKPPVVDDEAGSGLDNGDFKVTEDTST 240		
DB	181 KIKSSKNSAANRELQKLVKONKKNRTKKKPPVVDDEAGSGLDNGDFKVTEDTST 240		
QY	241 TQHNKVSTSPKITTAKPINRPSLPNSDTSEKTSITNKKETTVETKETTNNKOTSDG 300		
DB	241 TQHNKVSTSPKITTAKPINRPSLPNSDTSEKTSITNKKETTVETKETTNNKOTSDG 300		
QY	301 KETTSKAKETOSIEKTSADOLAPTSKVLAKPPKAEITTKGPALETTPKKEPTTPKBPAS 360		
DB	301 KETTSKAKETOSIEKTSADOLAPTSKVLAKPPKAEITTKGPALETTPKKEPTTPKBPAS 360		
QY	361 TPKEPTPTTIKSAPTTPKEBAPATTTKSAPTTPKEBAPATTTKSAPTTPKEBAPATTTK 420		
DB	361 TPKEPTPTTIKSAPTTPKEBAPATTTKSAPTTPKEBAPATTTKSAPTTPKEBAPATTTK 420		
QY	421 APTTKSAPTTPKEBAPATTPKKAPATTPKKEBAPATTPKKEBAPATTPKKEBAPATT 480		
DB	421 APTTKSAPTTPKEBAPATTPKKAPATTPKKEBAPATTPKKEBAPATTPKKEBAPATT 480		
QY	481 EPAPTPPKKAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKE 540		
DB	481 EPAPTPPKKAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKE 540		
QY	541 TTSAPTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTP 600		
DB	541 TTSAPTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTP 600		
QY	601 APTAPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKE 660		
DB	601 APTAPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKE 660		
QY	661 PEEBAPTPPAAAPNPKBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKE 720		
DB	661 PEEBAPTPPAAAPNPKBAPATTTPKEBAPATTTPKEBAPATTTPKEBAPATTTPKE 720		
QY	721 APTTPPKKAPKEAPATTTPKSTSDKAPATTTPKGAAPTTPKEBAPATTTPKEBAPATT 780		
DB	721 APTTPPKKAPKEAPATTTPKSTSDKAPATTTPKGAAPTTPKEBAPATTTPKEBAPATT 780		
QY	781 TAPTLKEBAPATTTPKAPKEAPATTTPKSTSDKAPATTTPKETAAPTTPKEBAPATTTP 840		
DB	781 TAPTLKEBAPATTTPKAPKEAPATTTPKSTSDKAPATTTPKETAAPTTPKEBAPATTTP 840		
QY	841 KPAPTPETPTPTST 900		
DB	841 KPAPTPETPTPTST 900		
QY	901 TKTPAATKPEMTTAKOKTERDLRTTETTTAAPKATKATATTEKTSKITATTTT 960		
DB	901 TKTPAATKPEMTTAKOKTERDLRTTETTTAAPKATKATATTEKTSKITATTTT 960		
QY	961 TSTTQDTTTPFKITTLKTTTLAATKVTTKTTTTEIMNKPESTAKPKDAPNASKATTP 1020		
DB	961 TSTTQDTTTPFKITTLKTTTLAATKVTTKTTTTEIMNKPESTAKPKDAPNASKATTP 1020		

THIS PAGE BLANK (USPTO)